Interactive Journal of Medical Research

Impact Factor (2022): 2.0
Volume 14 (2025) ISSN 1929-073X

Contents

Original Papers

Using an Electronic Immunization Registry (Aplikasi Sehat IndonesiaKu) in Indonesia: Cross-Sectional Study (e53849)
Dewi Aisyah, Astri Utami, Fauziah Rahman, Nathasya Adriani, Fiqi Fitransyah, M Endryantoro, Prima Hutapea, Gertrudis Tandy, Logan Manikam, Zisis Kozlakidis. 5

Glucagon-Like Peptide-1 Receptor Agonists Combined With Personalized Digital Health Care for the Treatment of Metabolic Syndrome in Adults With Obesity: Retrospective Observational Study (e63079)
Hala Zakaria, Hadoun Jabri, Sheikha Alshehhi, Milena Caccelli, Joelle Debs, Yousef Said, Joudy Kattan, Noah Almarzooqi, Ali Hashemi, Ihsan Almarzooqi. 177

Interactive Effects of Weight Recording Frequency and the Volume of Chat Communication With Health Care Professionals on Weight Loss in mHealth Interventions for Noncommunicable Diseases: Retrospective Observational Study (e65863)
Yuta Hagiwara, Takuji Adachi, Masashi Kanai, Kotoe Shimizu, Shinpei Ishida, Takahiro Miki. 189

Simplified Medication Adherence Questionnaire (SMAQ) for People Living With HIV in a National Hospital in Mexico: Instrument Validation Study (e59562)
Luis Del Moral Trinidad, Luz González Hernández, Jaime Andrade Villanueva, Pedro Martínez-Ayala, Adriana Valle Rodríguez, Vida Ruíz Herrera, José Vizcaíno Résendiz, Melva Herrera Godina, Sergio Dominguez-Lara. 212

Interstep Variations of Stairways and Associations of High-Contrast Striping and Fall-Related Events: Observational Study (e60622)
Sara Harper, Chayston Brown, Shandon Poulsen, Tyson Barrett, Christopher Dakin. 221

Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study (e49464)
Hurmat Shah, Mowafa Househ. 229

Student and Physician Views of How the Dobbs Decision Affects Training and Practice Location Preferences: Cross-Sectional Questionnaire Study (e55035)
Morgan Levy, Simone Bernstein, Sarah McNeilly, Abigail Liberty, Shira Fishbach, Shikha Jain, Jessica Gold, Vineet Arora. 243

Exploring Barriers to Patients' Progression in the Cardiac Rehabilitation Journey From Health Care Providers' Perspectives: Qualitative Study (e66164)
Shri Ramesh, Darwin Jull, Hélène Fournier, Fateme Rajabiyazdi. 255

Long-Term Engagement of Diverse Study Cohorts in Decentralized Research: Longitudinal Analysis of "All of Us" Research Program Data (e56803)
Vijay Yadav, Elias Neto, Megan Doerr, Abhishek Pratap, Larsson Omberg. 268

Policy Spotlight Effects on Critical Time-Sensitive Diseases: Nationwide Retrospective Cohort Study on Taiwan's Hospital Emergency Capability Categorization Policy (e54651)
Chih-Yuan Lin, Chih-Ching Liu, Yu-Tung Huang, Yue-Chune Lee. 281

Current Status and Future Directions of Ferroptosis Research in Breast Cancer: Bibliometric Analysis (e66286)
Jia-Yuan Luo, Yu-Long Deng, Shang-Yi Lu, Si-Yan Chen, Rong-Quan He, Di-Yuan Qin, Bang-Teng Chi, Gang Chen, Xia Yang, Wei Peng. . . 2 9 6

Evolution of Artificial Intelligence in Medical Education From 2000 to 2024: Bibliometric Analysis (e63775)
Rui Li, Tong Wu. 329

Increasing Participation and Completion Rates in Questionnaire Surveys of Primary Care Patients: Cluster-Randomized Study (e67981)
Paul Sebo, Benoit Tudrej, Augustin Bernard, Bruno Delaunay, Alexandra Dupuy, Claire Malavergne, Hubert Maisonneuve. 341

Correlation Between Objective Habit Metrics and Objective Medication Adherence: Retrospective Study of 15,818 Participants From Clinical Studies (e63987)
Antoine Pironet, L Phillips, Bernard Vrijens. 350

Perceptions and Experiences of Caregiver-Employees, Employers, and Health Care Professionals With Caregiver-Friendly Workplace Policy in Hong Kong: Thematic Analysis (e58528)
Maggie Lee, Eng-kiong Yeoh, Eliza Wong, Xue Bai, Nelson Yeung, Catherine French, Henock Taddese. 359

Evaluating the Effect of the JUUL2 System With 5 Flavors on Cigarette Smoking and Tobacco Product Use Behaviors Among Adults Who Smoke Cigarettes: 6-Week Actual Use Study (e60620)
Nicholas Goldenson, Saul Shiffman, Mark Sembower, Arielle Selya, Steve Pype, Ryan Black. 371

Associations Between Sleep Duration and Activity of Daily Living Disability Among Older Adults in China: Cross-Sectional Study (e65075)
Huimin Fan, Weijie Yu, Hongguo Rong, Xiaokun Geng. 392

Investigating the Association Between Mean Arterial Pressure on 28-Day Mortality Risk in Patients With Sepsis: Retrospective Cohort Study Based on the MIMIC-IV Database (e63291)
Qimin Chen, Wei Li, Ying Wang, Xianjun Chen, Dehua He, Ming Liu, Jia Yuan, Chuan Xiao, Qing Li, Lu Chen, Feng Shen. 404

Prevalence and Correlates of Clinically Elevated Depressive Symptoms in a Nationwide Sample of Transgender, Nonbinary, and Gender Diverse Young Adults in the United States: Cross-Sectional Survey Study (e66630)
Sari Reisner, Yuxin Liu, Regina Tham, Kaiden Kane, S Cole, Elizabeth Boskey, Sabra Katz-Wise, Alex Keuroghlian, Rena Xu. 416

Citizen Worry and Adherence in Response to Government Restrictions in Switzerland During the COVID-19 Pandemic: Repeated Cross-Sectional Online Surveys (e55636)
Vanessa Kraege, Céline Dumans-Louis, Céline Maglieri, Séverine Bochatay, Marie-Anne Durand, Antoine Garnier, Kevin Selby, Christian von Plessen. 428

French Versions of 4 English Questionnaires on Problematic Smartphone Use: Cross-Cultural Linguistic Translation and Adaptation Study (e53958)
Islam El Boudi, Mathilde Riant, Alexandre Bellier, Nicolas Vuillerme. 439

Viewpoints

How to Design Electronic Case Report Form (eCRF) Questions to Maximize Semantic Interoperability in Clinical Research (e51598)
Eugenia Rinaldi, Caroline Stellmach, Sylvia Thun. . . . 23

The Evolution of Uroflowmetry and Bladder Diary and the Emerging Trend of Using Home Devices From Hospital to Home (e66694)
Ming-wei Li, Yao-Chou Tsai, Stephen Yang, Yuan-Hung Pong, Yu-Ting Tsai, Vincent Tsai. . . . 30

The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations (e59823)
Dimitri Roustan, François Bastardot. . . . 38

An Automated Clinical Laboratory Decision Support System for Test Utilization, Medical Necessity Verification, and Payment Processing (e46007)
Safedin Beqaj, Rojeet Shrestha, Tim Hamill. . . . 45

Use of Clinical Public Databases in Hidradenitis Suppurativa Research (e70282)
Xu Liu, Linghong Guo, Xian Jiang. . . . 51

Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness (e60548)
Liesje Donkin, Nathan Henry, Amy Kercher, Mangor Pedersen, Holly Wilson, Amy Chan. . . . 55

Reviews

The Effect of Combining mHealth and Health Professional–Led Intervention for Improving Health-Related Outcomes in Chronic Diseases: Systematic Review and Meta-Analysis (e55835)
Masashi Kanai, Takahiro Miki, Toshiya Sakoda, Yuta Hagiwara. . . . 67

Comparing Digital Versus Face-to-Face Delivery of Systemic Psychotherapy Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e46441)
Pieter Erasmus, Moritz Borrmann, Jule Becker, Lars Kuchinke, Gunther Meinlschmidt. . . . 80

Global Evidence on the Sustainability of Telemedicine in Outpatient and Primary Care During the First 2 Years of the COVID-19 Pandemic: Scoping Review Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) Framework (e45367)
Daniela Valdes, Ankit Shanker, Ghofran Hijazi, Daniel Mensah, Tahir Bockarie, Ioana Lazar, Siti Ibrahim, Hamid Zolfagharinia, Rob Procter, Rachel Spencer, Jeremy Dale, Armina Paule, Liam Medlin, Keerthana Tharuvara Kallottil. . . . 101

Therapeutic Guidelines for the Self-Management of Major Depressive Disorder: Scoping Review (e63959)
Priscila de Campos Tibúrcio, Priscila Maria Marcheti, Daniela Miori Pascon, Marco Montebello Junior, Maria Alzete de Lima, Carla Fernandes, Célia Samarina Vilaça de Brito Santos, Maria Socorro de Sousa Nóbrega. . . . 123

Comparison of Outcomes Between Staged and Same-Day Circumferential Spinal Fusion for Adult Spinal Deformity: Systematic Review and Meta-Analysis (e67290)
Mert Dagli, Ryan Turlip, Felix Oettl, Mohamed Emara, Jaskeerat Gujral, Daksh Chauhan, Hasan Ahmad, Gabrielle Santangelo, Connor Wathen, Yohannes Ghenbot, John Arena, Joshua Golubovsky, Ben Gu, John Shin, Jang Yoon, Ali Ozturk, William Welch. . . . 143

Informatics Interventions for Maternal Morbidity: Scoping Review (e64826)
Jill Inderstrodt, Julia Stumpff, Rebecca Smollen, Shreya Sridhar, Sarah El-Azab, Opeyemi Ojo, Brendan Bowns, David Haggstrom. . . . 157

Integration of Conventional and Virtual Reality Approaches in Augmented Reality for Theory-Based Psychoeducational Intervention Design for Chronic Low Back Pain: Scoping Review (e59611)
Robin Conen, Steffen Mueller, Ana Tibubos. 200

Process Evaluations of Interventions for the Prevention of Type 2 Diabetes in Women With Gestational Diabetes Mellitus: Systematic Review (e51718)
Iklil Sa'id, Natasha Hotung, Madeleine Benton, Iliatha Nadal, Anisah Baharom, Matthew Prina, Barakatun Nisak Mohd Yusof, Kimberley Goldsmith, Samantha Birts, Ching Siew Mooi, Angus Forbes, Khalida Ismail, Boon Chew. 316

Original Paper

Using an Electronic Immunization Registry (Aplikasi Sehat IndonesiaKu) in Indonesia: Cross-Sectional Study

Dewi Nur Aisyah[1,2,3,4], PhD; Astri Utami[2], BSc; Fauziah Mauly Rahman[2,5], BSc; Nathasya Humaira Adriani[2], BSc; Fiqi Fitransyah[2], BSc; M Thoriqul Aziz Endryantoro[2], BSc; Prima Yosephine Hutapea[6], MPH; Gertrudis Tandy[6], MPH; Logan Manikam[1,3], PhD; Zisis Kozlakidis[7], PhD

[1]Department of Epidemiology and Public Health, Institute of Epidemiology and Health Care, University College London, London, United Kingdom

[2]Digital Transformation Office, Ministry of Health Republic of Indonesia, Jakarta, Indonesia

[3]Aceso Global Health Consultants Pte Limited, Singapore, Singapore

[4]Department of Public Health, Monash University, Tangerang, Indonesia

[5]School of Computer Science, Faculty of Engineering, University of Sydney, Sydney, Australia

[6]Directorate of Immunization Management, Ministry of Health Republic of Indonesia, Jakarta, Indonesia

[7]International Agency for Research on Cancer, World Health Organization, Lyon, France

Corresponding Author:
Logan Manikam, PhD
Department of Epidemiology and Public Health
Institute of Epidemiology and Health Care
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 2076792000
Email: logan.manikam.10@ucl.ac.uk

Abstract

Background: Electronic immunization registries (EIRs) are being increasingly used in low- and middle-income countries. In 2022, Indonesia's Ministry of Health introduced its first EIR, named Aplikasi Sehat IndonesiaKu (ASIK), as part of a comprehensive nationwide immunization program. This marked a conversion from traditional paper-based immunization reports to digital routine records encompassing a network of 10,000 primary health centers (*puskesmas*).

Objective: This paper provides an overview of the use of ASIK as the first EIR in Indonesia. It describes the coverage of the nationwide immunization program (Bulan Imunisasi Anak Nasional) using ASIK data and assesses the implementation challenges associated with the adoption of the EIR in the context of Indonesia.

Methods: Data were collected from primary care health workers' submitted reports using ASIK. The data were reported in real time, analyzed, and presented using a structured dashboard. Data on ASIK use were collected from the ASIK website. A quantitative assessment was conducted through a cross-sectional survey between September 2022 and October 2022. A set of questionnaires was used to collect feedback from ASIK users.

Results: A total of 93.5% (9708/10,382) of public health centers, 93.5% (6478/6928) of subdistricts, and 97.5% (501/514) of districts and cities in 34 provinces reported immunization data using ASIK. With >21 million data points recorded, the national coverage for immunization campaigns for measles-rubella; oral polio vaccine; inactivated polio vaccine; and diphtheria, pertussis, tetanus, hepatitis B, and *Haemophilus influenzae* type B vaccine were 50.1% (18,301,057/36,497,694), 36.2% (938,623/2,595,240), 30.7% (1,276,668/4,158,289), and 40.2% (1,371,104/3,407,900), respectively. The quantitative survey showed that, generally, users had a good understanding of ASIK as the EIR (650/809, 80.3%), 61.7% (489/793) of the users expressed that the user interface and user experience were overall good but could still be improved, 54% (422/781) of users expressed that the ASIK variable fit their needs yet could be improved further, and 59.1% (463/784) of users observed sporadic system interference. Challenges faced during the implementation of ASIK included a heavy workload burden for health workers, inadequate access to the internet at some places, system integration and readiness, and dual reporting using the paper-based format.

Conclusions: The EIR is beneficial and helpful for monitoring vaccination coverage. Implementation and adoption of ASIK as Indonesia's first EIR still faces challenges related to human resources and digital infrastructure as the country transitions from

paper-based reports to electronic or digital immunization reports. Continuous improvement, collaboration, and monitoring efforts are crucial to encourage the use of the EIR in Indonesia.

(Interact J Med Res 2025;14:e53849) doi:10.2196/53849

KEYWORDS

immunization; registry; digital; puskesmas; public health center; mobile app

Introduction

Background

Immunization is an important aspect of maintaining the health of individuals and society as a whole. It is a life-saving measure and a highly cost-effective public health intervention, an indispensable element of primary health care that shields children from avoidable mortality, illnesses, and disabilities caused by highly transmissible diseases [1]. Routine immunization plays an important role in reducing the incidence of various infectious diseases such as diphtheria, *Haemophilus influenzae* type B, measles, polio, rubella, and tetanus [2]. Located off the coast of Southeast Asia as the largest archipelagic country with 5 main islands, Indonesia has the fourth largest population in the world at >277 million [3]. This large population potentially poses the risk of rapid disease transmission if they are not protected by immunization. To ensure universal access to immunization services, vaccines must be distributed to geographically isolated regions as well as culturally or socially distinct populations, including hard-to-reach groups such as displaced individuals; migrants; and those impacted by conflict, political instability, and natural disasters [4]. Ensuring comprehensive immunization coverage for the entire population is of utmost importance yet challenging [3].

In recent years, Indonesia has faced challenges in achieving complete basic immunization, with 2020 and 2021 rates reaching 84.2%, falling short of the coverage targets of 92.9% and 93.6%, respectively [5,6]. Catch-up immunization for measles-rubella (MR) in toddlers also had declining coverage, standing at 65.3% (of a target of 76.4%) in 2020 and 58.5% (of a target of 81%) in 2021 [7]. Even with the increase in complete basic immunization coverage in 2022 (94.6%), immunization for preventable diseases remained low in low-coverage areas [5]. The COVID-19 pandemic significantly contributed to this decline as it led to disruptions in community immunization services, hindering the further attainment of herd immunity [5,6,8,9]. To address this issue, the Indonesian Ministry of Health (MoH) introduced the nationwide catch-up immunization campaign, which was called Bulan Imunisasi Anak Nasional (BIAN), in 2022 focusing on improving immunization coverage.

The MoH periodically conducts these nationwide catch-up immunization campaigns [7]. In 2022, Indonesia focused on conducting BIANs for two main activities: (1) supplementary immunization for MR and (2) catch-up immunization for children aged <5 years [7]. This focus was chosen based on several recommendations [7]. First, the National Committee for Measles-Rubella and Congenital Rubella Syndrome Elimination of Indonesia recommended to accelerate the achievement of MR and congenital rubella syndrome elimination targets in Indonesia by strengthening routine MR immunization with doses 1 and 2, aiming for a minimum coverage of 95%. Second, the National Immunization Technical Advisory Group recommended implementing inactivated polio vaccine (IPV) catch-up immunization for infants and children who had missed their scheduled IPV immunization to close immunity gaps and provide protection against the polio virus type 2. Third, the expert committee on diphtheria control recommended catch-up immunization efforts to close immunity gaps, especially among children aged <5 years (toddlers), and outbreak response immunization in areas experiencing outbreaks using appropriate vaccines for the target age group.

The BIAN was held initially for 1 month in May 2022 for the Sumatra, Kalimantan, Sulawesi, Nusa Tenggara, Maluku, and Papua islands and in August 2022 for the Java and Bali islands [7]. A total of 4 vaccines (oral polio vaccine [OPV]; IPV; MR; and diphtheria, pertussis, tetanus, hepatitis B, and *H influenzae* type B [DPT-HB-Hib]) were selected to support Indonesia's commitment to control global diseases, such as polio eradication, elimination of MR and congenital rubella syndrome and hepatitis B, control of diphtheria, reduction of the incidence of tuberculosis, and elimination of maternal and neonatal tetanus [7]. During the implementation of this BIAN, for the first time, Indonesia used an electronic immunization registry (EIR) capturing individual-level records [5] and replacing paper-based records as the latter were shown previously to include inaccuracies due to inexact, incomplete data and late reporting [10-12].

Objectives

EIRs as part of an immunization information system are a tool designed to provide information on immunization programs' target populations. By definition, EIRs include information that facilitates active search to identify individuals' vaccination data and vaccine history and provide support for determining which individuals need to be vaccinated and monitoring of immunization dropouts [13]. The 2 main databases needed for EIRs are demographic data and vaccination event data, which aim to identify the vaccine recipient and the vaccination event itself [13]. EIRs have been widely used in high-income countries as such digital reporting tools can improve reporting timeliness, precision, and overall performance [14]. Indonesia continues to improve its digital health sector and has launched the digital health transformation blueprint, aiming to create an integrated and sustainable health system [15]. As part of this effort, the government embarked on digitizing immunization reporting, shifting from a manual collection of aggregate data to individual digital records using the Aplikasi Sehat IndonesiaKu (ASIK). This paper aims to provide an overview of ASIK and its use and specifically evaluate ASIK implementation during the nationwide immunization campaign (BIAN) period in Indonesia.

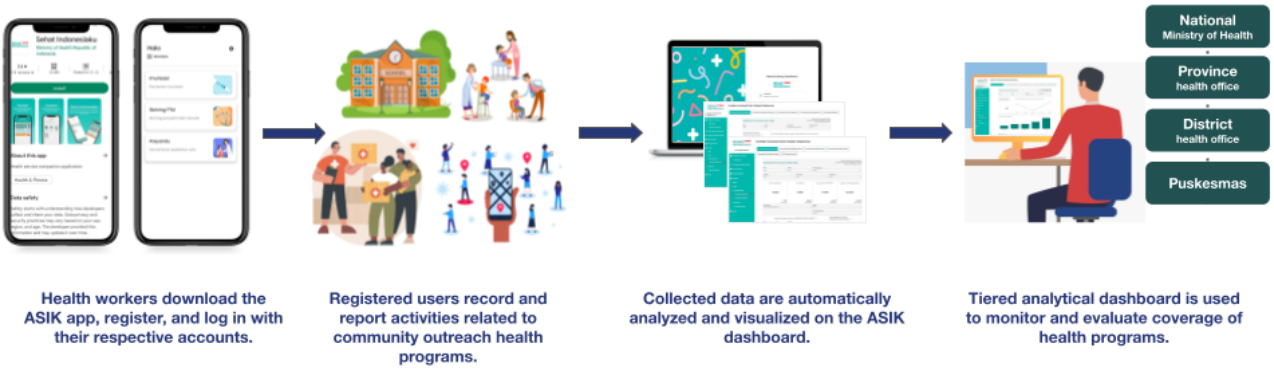
Methods

ASIK Overview

ASIK is an integrated application system developed by the MoH of Indonesia for health care workers to capture health care services provided for communities outside public health centers (*puskesmas*) [16,17]. This application is designed to help health care workers in primary care settings record patient data, report health care services provided for each individual, input routine health screening data, and monitor patient conditions [16,17].

The ASIK mobile app is available in an Android version and can be accessed through the Google Play Store. Targeted users of ASIK are health care workers in primary care settings. Health care workers need to register their account, as shown in Figure 1. Once registered and verified, health care workers can report any health care services provided outside the public health centers using ASIK to record individual data. Data reported through the ASIK mobile app are collected and automatically presented in an analytical dashboard provided in a web version of ASIK [17]. This web version of ASIK can only be accessed using a predetermined username and password for public health centers, district health offices, and provincial health offices to monitor and evaluate the coverage of health care services in their respective areas. The data presented in the dashboard are automatically synchronized from the public health centers level up to the national level. A simplified flow for the use of ASIK is shown in Figure 1.

Figure 1. Health service program reporting scheme using Aplikasi Sehat IndonesiaKu (ASIK). Puskesmas: Indonesian for "primary health center.".



ASIK was designed in line with elements of ideal EIRs according to the Pan American Health Organization's definition [18]. These elements encompass (1) inclusion of all persons at birth or as early as possible; (2) unique ID (national ID; biometrics or birth registration; or a unique combination of variables such as names, date or place of birth, and parental names or IDs); (3) information about each person, including information on geographical area of residence; (4) information about the vaccines administered, dates, and provider; (5) aggregation of data by geographical level as required; (6) timely individualized follow-up of vaccination schedule; (7) data entry as close to vaccination as possible (time and place); and (8) data security and protection of patient confidentiality.

The elements of ideal EIRs in ASIK were incorporated as follows: (1) allowing health workers to record individuals as soon as they are born with no restriction on age; (2) allowing health workers to record unique identification using the Indonesia resident registration number and a combination of name and date of birth as mandatory variables; (3) allowing health workers to record information about each individual using place of birth and current area of residence, which includes province name, city and district name, subdistrict name, village name, and detailed address; (4) allowing individuals to record information about vaccination, which includes vaccine name, vaccine batch number, date of vaccination, and location of vaccination; (5) a tiered dashboard that displays aggregated and structured data in spatial format from the national level to the village level; and (6) protection of patient confidentiality by ensuring a verification process during user registration and a limited access to a dashboard that displays detailed individual vaccination information (this access is designed specifically for the public health centers).

BIAN Immunization Reporting Using ASIK

The MoH launched and tested ASIK during the BIAN in May 2022 [19]. The BIAN 2022 immunization campaign was held in various mass public facilities such as *posyandu* (community-based health care service in Indonesia where local volunteers [cadres], midwives, and health workers provide essential maternal and child health services, including immunization); public health centers; subsidiary public health centers; hospitals; clinics; immunization service posts in schools, Islamic schools (*madrasas*), and Islamic boarding schools (*pesantrens*); and other strategic community-gathering places that could be turned into immunization posts [7]. Through ASIK, health care workers were able to record immunization data individually [16]. All immunization records from ASIK were monitored by the immunization program coordinators in public health centers in near real time.

Several databases served as the sources of ASIK data, including the public health center registry (Aplikasi Registrasi Puskesmas), the MoH Data and Information Center database, the Electronic Logistic Management and Information System (SMILE), and the Ministry of Home Affairs database. The Aplikasi Registrasi Puskesmas provides information regarding the location of immunization services, including names and locations of public health centers along with registered codes for each health facility. ASIK used data from the MoH Data and Information Center to acquire information on the number of targeted immunization participants for every district and city. Integration

with the SMILE database allowed ASIK to access data on the types of vaccines available, the quantity of vaccine doses, and the distribution locations. During the process of recording individual data, an interconnected application programming interface with the Ministry of Home Affairs ensured data validation for citizen identification numbers and domicile information.

Each record generated through ASIK is stored in a separate database and then automatically sent to the ASIK dashboard for data analysis. Processed data are distributed to various beneficiaries, such as health facilities, district and provincial health offices, and citizens, through different platforms. In addition, individuals who have been vaccinated can receive their immunization history data through the PeduliLindungi app, which was rebranded in 2023 as SATUSEHAT Mobile [20]. Figure 2 illustrates the entire data flow within ASIK.

The immunization data collection process in ASIK (Figure 3) comprises the steps outlined in Textbox 1.

Figure 2. Immunization data flow in Aplikasi Sehat IndonesiaKu (ASIK). OTP: one-time password; SMILE: Electronic Logistic Management and Information System.

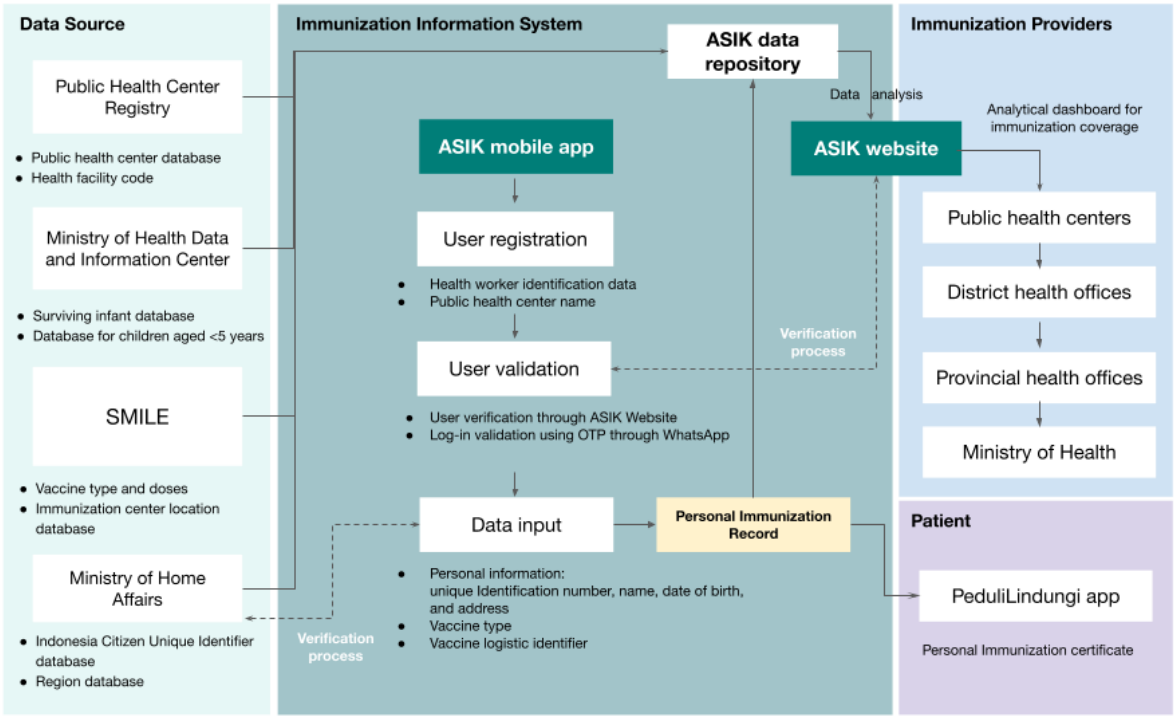
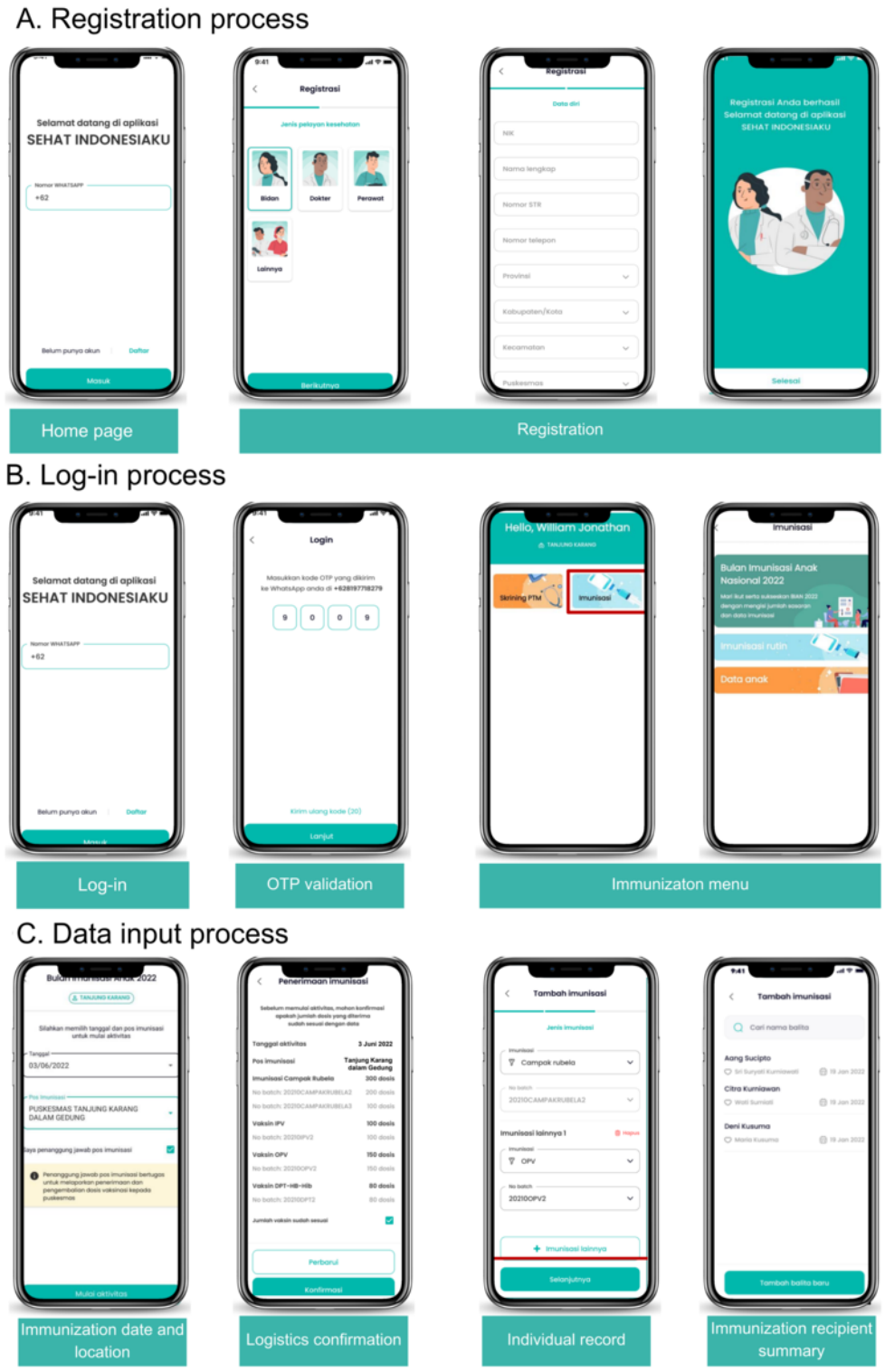


Figure 3. Aplikasi Sehat IndonesiaKu user interface for (A) registration process, (B) log-in process, and (C) data input process. OTP: one-time password.



Textbox 1. Immunization data collection process in Aplikasi Sehat IndonesiaKu.

User registration

Vaccinators are required to undergo registration to report immunization services using ASIK. The registration process involves providing personal information, specifying the work area, and selecting a specific public health center. Upon completing the registration form, users need to input their phone numbers to receive a one-time password for app log-in. The one-time password is sent through WhatsApp to each registered user to ensure system and user security.

Immunization data input

Vaccinators can record immunization data at designated immunization centers. Before selecting the immunization centers, the pharmaceutical officers must report the vaccine stock distribution to the specific immunization center through the SMILE. Once the logistic stock has been reported, vaccinators can choose a specific immunization center and the date of the immunization service. They can then input immunization data, including the names and addresses of recipients and the type of vaccine administered to each individual.

ASIK analytic dashboard

Immunization records captured by ASIK are integrated into an analytic dashboard. Public health center immunization coordinators can monitor the results of immunization services through this dashboard. The analytic dashboard also provides a spatial representation of the data, organized at provincial, district, and city levels. Thus, immunization data from the public health centers level to the national level can be monitored in near real time, with updates every 2 hours.

Individual immunization record

Parents can access their children's immunization records through the PeduliLindungi mobile app. Originally developed as a mass surveillance tool during the COVID-19 pandemic, PeduliLindungi initially presented COVID-19 vaccine certificates and later extended its functionality to include vaccine certificates from ASIK. Following the revocation of pandemic status, PeduliLindungi was further enhanced to encompass personal medical records. As a result, in 2023, the app was rebranded as a citizen health app named SATUSEHAT Mobile.

Data Collection

Data entry is initiated by selecting the BIAN from the immunization menu within ASIK. In total, 2 distinct roles can be chosen, namely, immunization program coordinator or immunization officer. The immunization program coordinator (ie, the individual in charge of the immunization posts) validates the available vaccine doses for each immunization post to activate the individual data report feature. This involves verifying data related to logistical distribution by reporting the number of vaccine distributions and the return of vaccination doses. On the other hand, immunization officers or vaccinators focus on reporting vaccination data at the individual level.

To record individual data, either the 16 digits of the national identification number or a combination of the full name and date of birth are entered to locate the appropriate personal health record. Consequently, a list of relevant individuals appears, allowing the users to select the relevant data or create a new identification record if the data are not available. Upon completing the identification data, vaccinators proceed to enter information about the administered vaccine type. To minimize error, the system was designed to display only the list of available vaccines suitable for the individual's age. For instance, for children aged 0 to 3 months, the system will not show the IPV in ASIK as the IPV has a minimum age requirement of 4 months. The system automatically records the results of the provided immunization services. It was recommended that vaccinators input the data immediately after services have been provided, although the option for backdating the data input is also available in ASIK, allowing users to submit delayed reports.

A quantitative survey was conducted from September 19, 2022, to October 21, 2022, using cross-sectional methods with a total of 1065 participants. All immunization program coordinators (ASIK users) from all public health centers, district health offices, and provincial health offices from 2 regions were invited to a web-based evaluation meeting. The first region represented a low data input into ASIK (the data gaps between manual and electronic reporting were approximately 30,000 to >100,000 immunization record data), and the second region represented the highest data input into ASIK (data gaps were of <10%). A total of 17.1% (182/1065) of the participants were provincial and district health officers, and 82.9% (883/1065) of the participants were health care workers from the public health centers. A questionnaire was used to collect feedback from the ASIK users (Multimedia Appendix 1). The survey used the Mentimeter platform, which enabled participants to provide direct answers and showed the cumulative, anonymized results directly. The questionnaire consisted of four sections: (1) overall user feedback on the system, (2) data reporting process, (3) supporting infrastructure, and (4) data completeness.

Data Analysis

Three outcomes were assessed in this study: (1) ASIK implementation coverage, (2) BIAN vaccination coverage (MR, OPV, IPV, and DPT-HB-Hib immunization), and (3) ASIK use feedback from users. ASIK implementation coverage and BIAN immunization coverage were calculated from the ASIK database. BIAN vaccination coverage was presented as percentage data; it was calculated by dividing the number of vaccinations (numerator) by the targeted population number for each vaccination (denominator) at the province level.

The results of BIAN immunization services were automatically presented in a structured dashboard that was accessible for public health centers, district health offices, provincial health offices, and the MoH. These data can be queried using predetermined variables. For instance, the analysis includes metrics such as the coverage of children administered the MR vaccine, OPV, IPV, and DPT-HB-Hib vaccine by comparing the service outcome rates (ie, the number of vaccinated children compared to the target numbers). The target estimates were

calculated based on the 2022 Health Development Program Target Population data [7]. Data analysis of the results of immunization services was also presented in a spatial format at the provincial, district, and city levels. Trends in service outcomes were depicted using time-series graphs. Moreover, comparative results of immunization coverage were generated (eg, by sex and age groups).

ASIK implementation coverage was measured by calculating how many public health centers used ASIK for the nationwide immunization program. The coverage was presented as percentage data, calculated by dividing the number of public health centers that reported BIAN immunization services using ASIK (numerator) by the total number of public health centers in Indonesia (denominator). The information was presented in the dashboard using the monitoring and evaluation feature. This feature enables relevant stakeholders to monitor the compliance of public health centers in reporting BIAN immunization services via ASIK. The monitoring and evaluation feature also provides information that represents the coverage of subdistricts, districts, and cities that have submitted their reports through ASIK. In addition, the dashboard generates a list of public health centers that have yet to submit their reports through ASIK. As a result, the provincial, district, and city health offices can effectively plan and implement a tiered monitoring and evaluation process.

ASIK use was assessed using a quantitative survey to obtain users' feedback. The outcome of the quantitative survey was to measure users' feedback on ASIK and how data collection using the digital tool was carried out, including (1) overall user feedback on the system (user understanding, user acceptance based on user interface [UI] and user experience [UX] and the variables presented, and user-friendliness of the system), (2) data reporting implementation (number of staff members for data input, flow of data reporting, and time to input data into ASIK), (3) infrastructure (mobile phone ownership and internet access), and (4) data completeness (availability of data on vaccinated individuals, comprising ID number, full name, date of birth, and gender). Outcomes 2, 3, and 4 were also explored to understand the real-life situations that potentially affected the use of an electronic reporting system in public health centers. The questions in the survey were developed as closed-ended questions, which allowed respondents to choose from a set of predefined answers quickly. The quantitative survey was analyzed using descriptive analysis. The data analysis was conducted using SPSS (version 20; IBM Corp) to describe users' feedback regarding the mobile app use in 2 different regions.

Ethical Considerations

The data collected for this paper do not require ethics approval in accordance with the Indonesian MoH *National Guideline and Standards for Ethical Research and Development in Health* guideline (2017, Chapter IIIB, point 2) [21] as no individual-level data are presented. The immunization analysis data are in anonymized, aggregate format and were handled following the aforementioned guidelines. For the survey, this research was conducted under the ethics approval of the International Agency for Research on Cancer Ethics Committee (22–37). Anonymous participation followed verbal informed consent, which was considered an indication of consent for obtaining and analyzing the data. There was no compensation provided for participants as it is part of the MoH immunization campaign program. No individual data were presented; only aggregated anonymized data were used as per the aforementioned guidelines.

Results

BIAN Immunization Reports by Province and Vaccine Type

The BIAN was held in 2 phases. The first phase began in May 2022 and ran until the end of July 2022, targeting 27 million children across 26 provinces in the Sumatra, Kalimantan, Sulawesi, Nusa Tenggara, Maluku, and Papua islands. The second phase took place in August 2022 in the Java and Bali islands (7 provinces), aiming for roughly 10 million children. The data analysis presented in this paper depicts the BIAN reports between May 2022 and December 2022. Different areas had distinct vaccination targets for each vaccine (Multimedia Appendix 2).

According to the data records in ASIK, there were 21.89 million BIAN immunization services recorded, with details on 18.3 million for the MR vaccine (18,301,057/36,497,694, 50.1%), 938,000 for the OPV (938,623/2,595,240, 36.2%), 1.2 million for the IPV (1,276,668/4,158,289, 30.7%), and 1.37 million for the DPT-HB-Hib vaccine (1,371,104/3,407,900, 40.2%).

The highest coverage of MR immunization was attained by Banten, reaching 96.8% (901,465/931,740), followed by Central Java (1,996,112/2,069,562, 96.5%) and East Java (1,797,337/2,352,409, 76.4%). In contrast, Aceh reported the lowest coverage (93,127/1,444,337, 6.4%), followed by Papua (72,368/792,523, 9.1%) and Central Sulawesi (104,316/708,642, 14.7%).

For OPV immunization, 2 provinces surpassed 100% coverage, namely, Lampung (49,833/37,827, 131.7%) and Banten (111,255/105,771, 105.2%), followed by Central Java (115,771/131,724, 87.9%). A significant contrast emerged in OPV coverage, with Papua (951/118,227, 0.8%), East Kalimantan (658/26,294, 2.5%), and Central Kalimantan (2857/58,389, 4.9%) showing the lowest coverage.

For IPV immunization, Banten reached the highest coverage (163,826/173,701, 94.3%), followed by Bali (1254/1426, 87.9%). In contrast, Papua (1814/167,678, 1.1%), East Kalimantan (747/67,544, 1.1%), and Central Kalimantan (1892/95,243, 2%) recorded the lowest IPV coverage.

For DPT-HB-Hib immunization, 2 provinces surpassed 100% coverage, namely, Banten (188,804/172,619, 109.4%) and Bali (3733/3620, 103.1%), followed by Lampung (64,043/71,504, 89.6%). Conversely, Papua (1294/194,510, 0.7%), Aceh (7282/184,475, 3.9%), and Central Sulawesi (2041/43,691, 4.7%) recorded the lowest coverage (Multimedia Appendix 3).

In summary, significant regional disparities in immunization coverage are apparent based on ASIK data. The Java and Bali islands exhibited the highest coverage across all vaccine types (MR: 7,544,342/9,435,097, 80%; OPV: 637,763/1,102,238,

57.9%; IPV: 927,925/1,747,414, 53.1%; DPT-HB-Hib: 1,087,268/1,759,356, 61.8%), whereas the Sulawesi island had the lowest coverage for 3 vaccine types (MR: 1,480,388/4,270,493, 34.7%; IPV: 25,270/406,208, 6.2%; DPT-HB-Hib: 28,745/315,809, 9.1%). Low coverage was also observed in the Papua, Maluku, and Nusa Tenggara islands for all vaccine types.

Using the specified BIAN 2022 campaign targets as benchmarks [7], we compared the immunization coverage achieved in each province with their predetermined targets. The immunization coverage report (Multimedia Appendix 4) illustrates that only 2 provinces, Banten and Lampung, successfully attained the coverage goal for MR immunization. In parallel, 4 provinces (Bali, Banten, Central Java, and Lampung) exceeded the target for OPV immunization. Similarly, the immunization target for IPV was met only by Bali, Banten, and Central Java. The DPT-HB-Hib immunization target was accomplished only by Bali, Banten, Central Java, and Lampung. These data are in line with past reports in which immunization targets were often unmet in the context of Indonesia, although the current technological implementation allowed for more granular data and for a quicker comparison at a national scale.

Digital Versus Paper-Based Recording of BIAN Implementation

Although the implementation of the BIAN and the release of ASIK happened simultaneously, the technical implementation guideline document for BIAN stated that manual or paper-based reports are still possible using a specified format [7]. Hence, all provinces involved also recorded data using a paper-based format. This manual data collection focuses on reporting aggregate numbers of immunization services, such as the total number of children aged 12 to 59 months receiving DPT-HB-Hib immunization. The data presented in the following paragraphs reflect the paper-based recording of the same campaign used to compare the digital versus paper-based recording of BIAN implementation.

Thus, according to the official letter from the MoH regarding the achievements of the BIAN phase I and phase II in 2022 and using the data from the paper-based records alone, the immunization coverage in Indonesia for MR, OPV, IPV, and DPT-HB-Hib was 72.7% (26,529,596/36,497,639), 54.2% (1,330,928/2,454,340), 45.8% (1,842,869/4,024,564), and 61% (2,011,057/3,294,942), respectively [22] (Multimedia Appendix 3).

The highest coverage of MR immunization was attained by East Java, reaching 100.6% (2,365,820/2,352,401), followed by Jakarta (710,757/715,786, 99.3%) and Banten (918,323/931,739, 98.6%). Conversely, Aceh reported the lowest coverage for MR (280,017/1,444,335, 19.4%), followed by Papua (307,466/792,523, 38.8%) and Riau (880,867/1,913,263, 46%).

In terms of OPV immunization, 3 provinces surpassed 100% coverage, namely, Bali (786/732, 107.4%), Central Java (113,325/106,194, 106.7%), and Banten (107,570/105,085, 102.4%). A significant contrast emerged in OPV coverage, with West Nusa Tenggara (10,224/129,413, 7.9%), West Kalimantan (6556/78,005, 8.4%), and Central Kalimantan (5472/58,389, 9.4%) showing the lowest coverage.

Regarding IPV immunization, Yogyakarta attained the highest coverage at 92.2% (400/434), followed by West Java (611,346/664,213, 92%) and East Java (195,047/212,128, 91.9%). Conversely, Papua (5521/167,678, 3.3%), East Kalimantan (2744/67,544, 4.1%), and Central Kalimantan (4378/95,243, 4.6%) recorded the lowest IPV coverage.

For DPT-HB-Hib immunization, South Sumatra led with the highest coverage (36,117/36,397, 99.2%), followed by Bali (3606/3639, 99.1%) and Central Java (293,073/297,231, 98.6%). In contrast, Papua (9676/194,510, 5%), Aceh (25,323/195,289, 13%), and West Kalimantan (8510/79,635, 10.7%) reported the lowest DPT-HB-Hib coverage (Multimedia Appendix 3).

On the basis of the table in the Multimedia Appendix 3, there were gaps between manual reporting data as compared to the digital recording via ASIK.

For MR immunization, the average reporting gap was 24.4% ±3.72% (95% CI: 20.68%-28.12%); the highest gap was observed in Maluku (47.4%), followed by South Sulawesi (47.3%) and Central Sulawesi (47.3%). The greatest reporting alignment was observed in Central Java (1.5%), followed by Banten (1.9%) and South Kalimantan (11.6%).

For OPV immunization, the average gap was 18.3% ±7.26% (95% CI: 11.04%-25.56%); the highest discrepancy was observed in East Java (70.4%), followed by Gorontalo (60.6%) and the Riau islands (53.6%). The greatest alignment was observed in Lampung (−39.7%), followed by Banten (−2.8%) and Jakarta (−1.9%).

For IPV immunization, the average gap was 16.1% ±6.42% (95% CI: 9.68%-22.52%); the highest data discrepancy was observed in the Special Region of Yogyakarta (77.4%), followed by East Java (70.8%) and the Riau islands (56.5%). The closest reporting alignment was observed in Jakarta (−1.8%), followed by West Nusa Tenggara (1.6%) and Papua (2.2%).

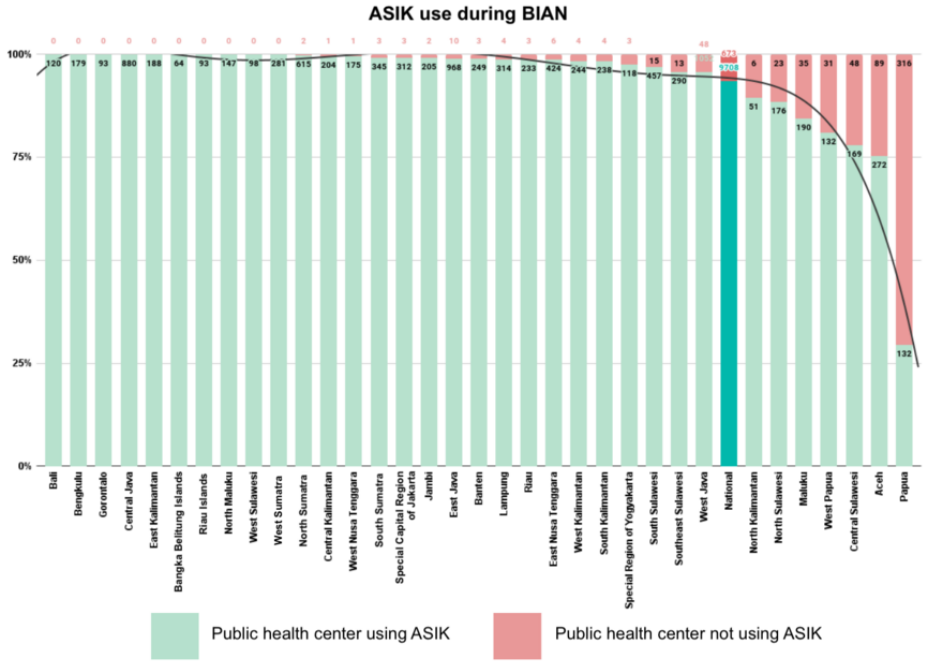
For DPT-HB-Hib immunization, the average gap was 20.7% ±7.23% (95% CI: 13.47%-27.93%); the highest data reporting discrepancy was observed in the Riau islands (82.4%), followed by East Java (71.6%) and Maluku (58.2%). The greatest reporting alignment was observed in Banten (−12.2%), followed by Bali (−4%) and Lampung (−0.4%).

ASIK Use and Coverage

To understand the aforementioned discrepancies, 2 further aspects were considered: ASIK use and coverage and user feedback on ASIK use. The results will be presented in the following paragraphs. Regarding ASIK coverage, a total of 93.5% (9708/10,382) of public health centers, 93.5% (6478/6928) of subdistricts, and 97.5% (501/514) of districts and cities in 34 provinces reported BIAN activities using ASIK. On the basis of the coverage as measured at the province level, there were 10 provinces where 100% of their public health centers reported BIAN activities using ASIK, namely, Bali, Bengkulu, Gorontalo, Central Java, East Kalimantan, Bangka Belitung Islands, Riau Islands, North Maluku, West Sulawesi, and West Sumatra, whereas the provinces with the lowest

number of public health centers reporting using ASIK were Papua (132/448, 29.5%), Aceh (272/361, 75.3%), and Central Sulawesi (169/217, 77.9%). Detailed information on other provinces can be found in Figure 4.

Figure 4. Aplikasi Sehat IndonesiaKu (ASIK) use across 34 provinces during the Bulan Imunisasi Anak Nasional (BIAN) implementation.



User Feedback on ASIK Use

In parallel to understanding the coverage and use of ASIK, a quantitative survey was conducted to obtain users' feedback on the system, including explorations of UI and UX, reporting variables, and system interference. A total of 1065 questionnaires were received, although not all respondents answered all questions. The results of the quantitative survey are presented in this section and summarized in Table 1. In general, users exhibited a good understanding of ASIK as the EIR (650/809, 80.3%). For UI and UX, the highest feedback was that it was overall good but could still be improved (489/793, 61.7%), with 35.3% (280/793) of respondents expressing that the UI and UX were very good and 3% (24/793) saying that they were hard to understand. Regarding the required variables for immunization data input in ASIK, 54% (422/781) of the respondents expressed that the variables fit their needs but could still be improved, 39.6% (309/781) expressed that the variables fit their needs, and 6.4% (50/781) of users expressed that there were too many variables. In terms of system interferences, 59.1% (463/784) of users expressed that they sometimes happened, and 24.5% (192/784) expressed that they often happened. The most prevalent system interferences were (1) inability to find individual data in the system (281/779, 36.1%), (2) server errors (217/779, 27.9%), and (3) problems with logistic stock confirmation (135/779, 17.3%).

On the basis of the users' feedback, 44% (382/869) stated that there were 2 to 3 staff members that helped with the data reporting process. However, the number of staff members who helped with data input varied as users in region 2 (high data collection) had >5 staff members for data reporting and users in region 1 (low data collection) only had 2 to 3 staff members to do so. Regarding the flow of data reporting, 33.4% (291/872) of users exhibited a combination of recording the immunization data on paper or on Microsoft Excel first and directly inputting them into ASIK. Furthermore, more users preferred the input of immunization data at public health centers right after the immunization activity was completed (364/1089, 33.4%), followed by inputting the data within 24 hours after the immunization services (203/1089, 18.6%). However, 22.3% (181/813) of users in region 1 input the immunization data 2 weeks after the immunization activity, whereas in region 2, a total of 45.3% (125/276) of their users preferred to input the data right after the activity was completed. Regarding the data reporting process, each region had different characteristics for each variable: number of staff members, data reporting flow, and time to input the data. Region 2 showed a relatively higher number of staff members, more direct flow of data reporting, and timelier input of data into the system. As previously mentioned, region 2 had the highest data collection, comprising 28 cities that did not have any gaps in data collection between ASIK and manual reporting.

Table 1. Quantitative exploration of the use of Aplikasi Sehat IndonesiaKu (ASIK; N=1065).

Characteristic	Region 1/ (n=800), n/ (%)	Region 2/ (n=265), n/ (%)	Total, n/ (%)
Overall user feedback on the system			
System understanding			
Limited understanding	135/598 (22.6)	24/211 (11.4)	159/809 (19.7)
Good understanding	463/598 (77.4)	187/211 (88.6)	650/809 (80.3)
User interface and user experience			
Very good	200/587 (34.1)	80/206 (38.8)	280/793 (35.3)
Overall good but can still be improved	366/587 (62.4)	123/206 (59.7)	489/793 (61.7)
Hard to understand	21/587 (3.6)	3/206 (1.5)	24/793 (3)
Variables			
Fit with needs	213/576 (37)	96/205 (46.8)	309/781 (39.6)
Fit with needs but can still be improved	321/576 (55.7)	101/205 (49.3)	422/781 (54)
Too many variables	42/576 (7.3)	8/205 (3.9)	50/781 (6.4)
System interference			
Seldom	18/579 (3.1)	3/205 (1.5)	21/784 (2.7)
Rarely happened	58/579 (10)	36/205 (17.6)	94/784 (12)
Sometimes happened	326/579 (56.3)	137/205 (66.8)	463/784 (59.1)
Often happened	163/579 (28.2)	29/205 (14.1)	192/784 (24.5)
Always happened	14/579 (2.4)	0/205 (0)	14/784 (1.8)
Data reporting process			
Number of staff members for data input			
1 person	95/645 (14.7)	8/224 (3.6)	103/869 (11.9)
2-3 people	311/645 (48.2)	71/224 (31.7)	382/869 (44)
4-5 people	98/645 (15.2)	43/224 (19.2)	141/869 (16.2)
>5 people	141/645 (21.9)	102/224 (45.5)	243/869 (28)
Flow of data reporting			
(1) Recording the data on paper or manual book report before inputting them into ASIK	153/647 (23.6)	35/225 (15.6)	188/872 (21.6)
(2) Recording the data in Microsoft Excel sheet before inputting them into ASIK	119/647 (18.4)	20/225 (8.9)	139/872 (15.9)
(3) Directly inputting the data into ASIK	51/647 (7.9)	46/225 (20.4)	97/872 (11.1)
Combination of 1 and 2	131/647 (20.2)	26/225 (11.6)	157/872 (18)
Combination of 1, 2, and 3	193/647 (29.8)	98/225 (43.6)	291/872 (33.4)
Time to input data into ASIK			
Directly at the vaccination center	78/813 (9.6)	60/276 (21.7)	138/1089 (12.7)
Directly after the activity at the public health centers	239/813 (29.4)	125/276 (45.3)	364/1089 (33.4)
Input the data within a day at home	161/813 (19.8)	42/276 (15.2)	203/1089 (18.6)
Input the data within 3 days of manual data collection	86/813 (10.6)	14/276 (5.1)	100/1089 (9.2)
Input the data within 7 days of manual data collection	68/813 (8.4)	14/276 (5.1)	82/1089 (7.5)
Input the data ≥2 weeks after manual data collection	181/813 (22.3)	21/276 (7.6)	202/1089 (18.5)
Existing infrastructure			
Android mobile phone ownership			
Yes	694/711 (97.6)	229/241 (95)	923/952 (97)[y]

Characteristic	Region 1/ (n=800), n/ (%)	Region 2/ (n=265), n/ (%)	Total, n/ (%)
No	17/711 (2.4)	12/241 (5)	29/952 (3)[y]
Internet access			
Available at all places	390/710 (54.9)	163/240 (67.9)	553/950 (58.2)
Only available at the public health centers	78/710 (11)	14/240 (5.8)	92/950 (9.7)
Limited availability	242/710 (34.1)	63/240 (26.3)	305/950 (32.1)
Data completeness			
Data availability			
Individual data were complete	157/637 (24.6)	125/219 (57.1)	282/856 (32.9)
Individual data were incomplete	480/637 (75.4)	94/219 (42.9)	574/856 (67.1)

Regarding the existing infrastructure, 97% (923/952) of users had Android mobile phones. A total of 58.2% (553/950) of respondents expressed that internet access was available at all places, 32.1% (305/950) expressed that internet availability was limited in several places, and 9.7% (92/950) expressed that the internet was only available at public health centers (and not available in immunization posts). The data suggested that the internet was more accessible in region 2 compared to region 1.

Another constraint found during data reporting was data completeness of the targeted participants. A total of 67.1% (574/856) of users expressed that the individual data were incomplete, especially on unique identifier (ID number), which was only available for up to 79% of the collected data points; full name, which was available for up to 95% of individual data points; birth date, which was available for up to 95% of the collected data points; and gender, which was available for up to 95% of the collected data points.

Most respondents (618/1065, 58%) indicated several constraints that hindered the use of ASIK and were experienced intermittently or too frequently, including (1) challenges in retrieving individual data, (2) server errors, and (3) logistic stock validation. The first challenge that arose in the early development of the system is the lack of interconnectedness with the civil registration database. During the preliminary development phase of ASIK, the absence of established data sharing agreements between the Indonesian MoH and the Ministry of Home Affairs prevented users from efficiently engaging in data reporting using ASIK.

Interconnection between ASIK and the civil registration database will help health care workers automatically retrieve individual demographic information such as gender, date of birth, and address by inputting the 16 digits of the national ID. During BIAN implementation, ASIK was still not connected to the civil registration database owned by the Ministry of Home Affairs; thus, health care workers needed to input the individual information manually.

The second challenge also stemmed from the early development phase of the system, when the initial server capacity was constrained, leading to suboptimal application performance. System errors that were mentioned by the respondents happened occasionally throughout the recording process. They sometimes happened due to personal data input, vaccine type data input, or the process of saving data. They potentially happened due to limited server performance caused by the high number of users inputting data at the same time.

The third challenge pertains to logistic stock management. In contrast to the first challenge, ASIK was strategically integrated with the logistic system (SMILE) during its development phase, effectively establishing a linkage between immunization and logistic data. However, the compliance and adherence to using the SMILE were variable. Consequently, the comprehensive implementation of the SMILE for monitoring logistic stocks in the context of the BIAN 2022 campaign was not universal across all public health centers. This posed an obstacle in the use of ASIK given that logistic confirmation is a prerequisite for inputting immunization data into ASIK. Respondents were hindered from inputting the immunization data into ASIK if the pharmacy staff had not input the logistic information into the SMILE. This created some delays in the data recording process during BIAN 2022.

These challenges have been addressed, and the system has been enhanced over time, particularly during the latter stage of the BIAN 2022 campaign. The system was integrated with the civil registration database, the server capacity was upgraded, and refinements were implemented in the integration between ASIK and the SMILE. As a result, the efficacy of ASIK had a notable improvement during the latter phase of the BIAN 2022 campaign. Nevertheless, many respondents gave positive responses for the use of ASIK as the Indonesia EIR. The positive response correlated with the convenience of ASIK as an EIR.

Discussion

Principal Findings

The launch of ASIK is a part of the MoH's mission to transform health care services by assisting health care workers in efficiently and comprehensively recording patient data in a single, integrated database [5,15,19]. The use of ASIK as Indonesia's first EIR was broad, covering 34 out of 38 provinces. The rate of adoption reached 93.5% (9708/10,382) of public health centers (ie, the main health care provider for the nationwide immunization campaign). Through 2 phased releases, ASIK managed to record >21 million immunizations. Although most users expressed a good understanding of the

system, several challenges persisted during the implementation phase.

EIRs are instruments proven to be cost-effective in enhancing coverage and timeliness of vaccination [23]. The use of EIRs was considered beneficial in preventing delays in vaccinating children as they help parents keep track of their children's vaccination records and stay informed of their children's vaccination schedule [23,24]. Furthermore, EIRs are helpful to identify high-risk populations; facilitate resource and activity planning, such as monitoring vaccination coverage and vaccine stock availability; and determine the overall performance of the immunization program [23,24]. During ASIK's first launch, Indonesia's MoH emphasized that, in the future, children who have received immunization will have their records digitally stored in ASIK, in which the integrated data aim to facilitate parents' access to their children's immunization data, especially to be used for school-related purposes [19].

In the technical implementation guideline document for the BIAN, recording and reporting of immunization activities are mandated to be conducted electronically through ASIK [7]. Despite electronic data input being mandatory, the data recorded in ASIK did not represent the complete immunization service during the BIAN 2022 campaign. According to the ASIK analytical dashboard, the national coverage recorded in ASIK for MR, OPV, IPV, and DPT-HB-Hib immunization was 50.1% (18,301,057/36,497,694), 36.2% (938,623/2,595,240), 30.7% (1,276,668/4,158,289), and 40.2% (1,371,104/3,407,900), respectively. In addition, there was an average discrepancy of approximately 19% between digital and manual data reporting. When compared to similar supplementary immunization efforts in previous years, the Polio National Immunization Week held in 2016 achieved an impressive coverage rate of 96.5%. MR supplementary immunization activities in 2017 exceeded expectations, reaching >100.9%, whereas in 2018, the coverage stood at approximately 73.4% [7]. Around the same time, between 2017 and 2019, several countries, including Afghanistan, Benin, Ghana, Timor-Leste, Togo, Sierra Leone, Senegal, Rwanda, Pakistan, Niger, Malawi, Lesotho, and Ethiopia, also attained MR supplementary activity coverage rates of >90% [25]. Nonetheless, it is important to consider that these coverage data were obtained from manual or paper-based immunization records and postcampaign surveys [25].

Considering that the immunization data are stored in different files segregated by health center and area, it is nearly impossible to ensure that there is no duplication of data in each record. Studies show that paper-based data reports actually create challenges such as affecting health care workers' ability to make data-driven decisions by analyzing the manual immunization reports and increasing the burden of health workers to input data both in the paper-based and digital reporting systems [26]. Furthermore, the use of paper-based aggregate data for analysis presents gaps, such as limited individual longitudinal data, and omits crucial sociodemographic information [27]. Data stored electronically enable prompt data retrieval, hence facilitating immediate reports related to disease and surveillance, which is in contrast to paper-based reports that rely on manual retrieval [13]. Dual use of EIR and paper-based immunization reports further exacerbates the challenges of data input using ASIK for health care workers. A transition from paper-based reports to paperless reports and dual entry processes are also perceived barriers and challenges to EIR implementation in other countries such as Kenya and Vietnam [26,28]. The immunization program has grown in complexity and leads to an increased need for data, and the immunization information system must be flexible to cater to those needs [13]. Paper-based reports are limited in meeting such needs as they are heavily influenced by population mobility and errors stemming from estimation of population number [13]. Hence, moving forward, it is important for relevant stakeholders to decide whether to solely use paper-based reports or onboard using ASIK for Indonesia immunization programs.

On the other hand, data captured in ASIK represent individual data, which can be linked to personal immunization history records. Individual immunization data recorded in an EIR such as ASIK offer the advantage of providing precise and timely information. This, in turn, facilitates critical activities such as monitoring vaccination compliance, mitigating dosing errors, and identifying unvaccinated individuals [29]. The significance of individual immunization records in clinical decision-making is noteworthy as they provide a comprehensive historical vaccination record and diminish the likelihood of misclassifying vaccination status [30]. Consequently, this aids in reducing instances of both under- and overimmunization in children [31]. Furthermore, integrating individual immunization data from an immunization information system into an electronic medical record or electronic health record can enhance the accuracy of vaccine safety assessments [30,32].

ASIK has achieved significant engagement, involving 93.5% (9708/10,382) of public health centers across Indonesia within a span of 7 months since its launch. This indicates that public health centers were informed about the data input process and user guide updates that come with it, although long-term adherence to data input remains to be seen. The high level of ASIK use might be supported by several aspects, such as the compatibility of the Android app in Indonesia, the user-friendliness of the system interface, and the ease of use. Considering the high level of mobile phone and internet use in Indonesia, the digitization of health service reports is happening apace [33]. In 2021, approximately 90.5% of households in Indonesia owned at least one mobile phone, covering 65.8% of the population, and approximately 62% of the population had access to the internet [34]. Indonesia is estimated to have >200 million mobile phone users and is dominated by Android operating systems (95%) [34,35]. This is in line with our findings, in which 97% (923/952) of respondents stated that they owned Android-based mobile devices. ASIK is built on an Android operating system, which is compatible with various mobile devices, allowing health care workers in Indonesia to benefit from it [33]. Existing EIRs such as ImmunizaCA in Canada, Zindagi Mehfooz in Pakistan, and DHIS2 Android Capture App in Zambia are also built on Android-based platforms [36-40].

A visually appealing or user-friendly UI and UX can prompt users to explore an app, facilitating a comprehensive understanding of its available functions, consequently influencing user acceptability [41,42]. It is essential to acknowledge that ASIK signifies a transition for health care

workers from paper-based to electronic reporting; hence, data and variables should be both useful and recognizable to the health workers as it is important to note that a familiar interface correlates with use frequency [43]. Other than that, a familiarity with workflow and a design that carefully considers health care workers' needs are crucial for effective EIR implementation [26].

Despite the users' acceptability of ASIK as an EIR, the low number of reporting in certain provinces may be due to several factors. First, primary health care workers, often responsible for numerous data collection tasks, are frequently burdened with heavy workloads, lack motivation, and view data recording as undesirable and time-consuming [44,45]. Such excessive workloads can compromise the quality of health care data reported at health facilities, leading to inaccuracies and unreliability when other tasks interfere with data entry [46,47]. Indeed, the lowest compliance was often correlated with the provinces that have the lowest ratios of primary health care staff, and thus, it is presumed that staff time availability is the strongest root cause for the observed variation in ASIK use. The workload also affects workers' opportunities to adequately familiarize themselves (digital familiarization) with any health information system they are about to use, often coupled with inadequate handover and shortage of available staff for this knowledge transfer [48,49]. In the next phase, to ensure that there is an appropriate and sustainable knowledge transfer to all end users, conducting a cascading training-of-trainers approach and establishing a multilevel technical support network is recommended [28].

The sudden explosion in the number of health-related apps in Indonesia, with >400 health-related apps developed and introduced for health care facilities, further exacerbates the burden on health workers regarding digital familiarization and data input [15]. In addition, lack of adequate monitoring by supervisors or managers and provision of technical support affects the effective implementation of data reports through an information system [48,49]. Furthermore, the BIAN implementation coincided with a massive COVID-19 preventive immunization effort, adding more tasks to primary care workers' daily workload [50,51]. The implementation of COVID-19 immunization entailed a collaborative approach involving various sectors, with health personnel sourced from governmental, military, and private practitioner sectors [52]. This stands in stark contrast to the BIAN 2022 campaign, which relied heavily on health workers stationed at public health center facilities [52]. Therefore, there was a double burden for primary care workers to reach both the target for COVID-19 immunization and the target for the BIAN 2022 campaign and report both results using different digital pathways. This further justified the need to simplify, integrate, and ensure the interoperability of the health information system in Indonesia, as mentioned in the digital transformation blueprint, and ensure that the EIR is developed to match users' needs and able to simplify the reporting workflow accordingly [15,26].

In the process of implementing digital technologies in the health care sector, certain challenges are inevitable [53]. These challenges encompass a range of factors, including but not limited to client identification and verification as well as lack of database integration [53,54]. In ASIK, the collection of the national identification number of the service recipient is a prerequisite for generating an immunization service record. Indonesia uses a 16-digit identification number, known as the resident registration number, which serves as a cornerstone of national demographic identification [55]. Current data indicate that 97.2% of Indonesian citizens already possess a unique identification number, but there are also 5.38 million individuals without unique identification numbers [55,56]. During BIAN implementation, there were many instances in which patients did not bring any personal identification tools when immunization services were provided. Consequently, users were burdened to create individual data without unique identifiers while also needing to remember that exact record for identification and retrieval in the future when the same patient visited. Lack of linkage and integration with other databases were also a common challenge, which happened in the development of EIRs in Africa [54]. It is crucial to foster awareness and establish a mutual understanding between the realms of public health and IT to facilitate EIR development [54]. Ensuring that EIRs comprehensively covered all regions across Indonesia, encompassing an estimated target population of 37 million individuals for BIAN immunization, presented a challenge in conveying this requirement effectively, especially in the beginning of the system's development. To build an interoperable EIR system, integration with the national ID database is crucial [28]. Moving forward, the MoH could establish a continuous collaboration with the Ministry of Home Affairs as the primary stakeholders and foster understanding of and commitment to the need for database integration and providing immediate legal identity to the population since birth.

To report immunization data in ASIK, having access to the internet is a prerequisite. Despite statistics indicating that approximately 62% of Indonesia's population use the internet and the country's consistent progress in improving internet connectivity, disparities persist in terms of access across different social classes and regions [34,57]. Notably, significant portions of Indonesia, particularly in the eastern regions, still lack access to the internet [58]. The distribution of infrastructure for signal reception in Indonesia reveals that >3023 villages lack signal reception infrastructure, with >934 of these villages located in Papua [34]. Given the limited internet access, it is unsurprising that the coverage of MR, OPV, IPV, and DPT-HB-Hib immunization reports using ASIK remained low in Papua. Mobile apps that need to constantly rely on internet connectivity reduce effectiveness for users such as health care workers that operate in remote regions [59]. To reduce the digital divide among regions, the government could explore the potential use of satellite-based cellular and internet connectivity for rural areas [60]. This could be done through collaboration with the Ministry of Communication and Information Technology or alongside existing internet provider entities. Other than ensuring internet connectivity, support and assistance to infrastructure and human resources is paramount, particularly in areas with low information and communications technology maturity scores [61].

This is the first study that describes the EIR in Indonesia, which replaces the manual reporting system from the MoH, including

describing its use during the BIAN. Furthermore, the quantitative survey provided an overview of the obstacles in the implementation of ASIK for the BIAN program and provided input on the features that need to be improved for future immunization data collection processes. However, a relative weighing of the factors that require future improvement was not sought and, thus, can become the subject of future research.

This study has several limitations that need to be considered. First, ASIK is only available on Android-based platforms and has not been released for iOS; thus, data reporting is limited to officers who own devices compatible with Android. Second, the quantitative survey that was administered covered a limited number of the actual ASIK users in the field, but it is expected that this number can represent feedback from the field. In addition, using Mentimeter potentially limited the feedback from some of the participants when they did not fill the questionnaire in a timely manner, and considering that the evaluation meeting was conducted on the web, there was a risk of limited control over the participants filling out the questionnaire in a timely manner. Finally, the manual data that were compared to ASIK are the result of a report from the provincial health office, which might not fully capture the recording process that took place. Nevertheless, this study provides a deeper understanding of the use of ASIK in the BIAN program as the first EIR implemented nationally by the MoH.

Conclusions

EIRs have demonstrated their cost-effectiveness in improving vaccination coverage and timeliness, offering numerous advantages in disease prevention and health care management. Adoption and continuous use of EIRs entail numerous challenges, especially in a country such as Indonesia with an enormous number of targeted immunization populations and limited infrastructure settings. Challenges observed from this study included the gap between manually reported data and electronic records due to dual data entry processes; heavy workloads for primary care health workers to carry out immunization campaigns as well as the reporting process; lack of data completeness in ID number, full name, date of birth, or gender; system interferences due to the early development phase of the system; and limited infrastructure settings to support digital connection in remote areas. Addressing these challenges and ensuring adequate support for health care workers are essential steps in enhancing the effectiveness of EIRs in improving public health outcomes.

Although faced with the aforementioned challenges, ASIK as Indonesia's first national EIR managed to reach substantial engagement. More than 93% of its targeted users adopted the system, and this was potentially made possible by ease of system use, continuous system improvement according to users' feedback, and immediate regulation support. Ease of use and user-friendly apps that simplify data input processes are essential to reinforce user acceptability. In the future, ASIK can be developed for a wider immunization program provided that improvement of features is carried out, including the preparation of relevant human resources and supporting digital infrastructure and stronger regulatory support.

Acknowledgments

The authors would like to thank the Digital Transformation Office, Ministry of Health, Republic of Indonesia, for supporting the study implementation and developing the Aplikasi Sehat IndonesiaKu integrated system; the public health centers and district and provincial health offices across Indonesia who participated in the study; and the Directorate of Immunization Management, Ministry of Health of Indonesia, who supported the study implementation. The authors did not receive funding for the creation of this manuscript. This work was supported in part by the International Agency for Research on Cancer (IARC) through training by the Biobanking and Cohort Network (Civilian Research and Development Foundation Global, award 66415). While 2 authors are affiliated with Aceso Global Health Consultants, which is a private company, they declare that this research project did not receive funding from Aceso Global Health Consultants. The company did not have a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Where authors are identified as personnel of the IARC, the authors alone are responsible for the views expressed in this paper, which do not necessarily represent the decisions, policy, or views of the IARC.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DNA led the study design and developed the paper concept, main structure, and methodology. AU, FMR, and NHA conducted the quantitative survey and accessed and analyzed the data. DNA, AU, FMR, and NHA wrote the first draft of the paper. ZK and LM validated the study and critically revised the manuscript content. DNA and AU had primary responsibility for the final content. All authors contributed to reviewing the study data and the manuscript design and to the approval of the final manuscript.

Conflicts of Interest

LM is the director of Aceso Global Health Consultants, and DNA is a consultant of the company. However, both of them contributed to this paper on a pro bono basis. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 203 KB - ijmr_v14i1e53849_app1.docx]

Multimedia Appendix 2
Bulan Imunisasi Anak Nasional 2022 campaign target.
[DOCX File , 9 KB - ijmr_v14i1e53849_app2.docx]

Multimedia Appendix 3
Comparison of measles-rubella vaccine; oral polio vaccine; inactivated polio vaccine; and diphtheria, pertussis, tetanus, hepatitis B, and Haemophilus influenzae type B immunization coverage using Aplikasi Sehat IndonesiaKu data and manual report data.
[DOCX File , 14 KB - ijmr_v14i1e53849_app3.docx]

Multimedia Appendix 4
Detail of measles-rubella vaccine; oral polio vaccine; inactivated polio vaccine; and diphtheria, pertussis, tetanus, hepatitis B, and Haemophilus influenzae type immunization coverage using Aplikasi Sehat IndonesiaKuASIK data and manual report data.
[DOCX File , 23 KB - ijmr_v14i1e53849_app4.docx]

References

1. Vaccine-preventable diseases and immunization programme. World Health Organization Regional Office for Europe. 2005. URL: https://iris.who.int/bitstream/handle/10665/107665/WHO-EURO-2005-8588-48360-71799-eng.pdf?sequence=3&isAllowed=y [accessed 2023-08-05]
2. Talbird SE, Carrico J, La EM, Carias C, Marshall GS, Roberts CS, et al. Impact of routine childhood immunization in reducing vaccine-preventable diseases in the United States. Pediatrics 2022 Sep 01;150(3):e2021056013. [doi: 10.1542/peds.2021-056013] [Medline: 35821599]
3. Indonesia population. Worldmeter. URL: https://www.worldometers.info/world-population/indonesia-population [accessed 2023-08-05]
4. Immunization Agenda 2030 Partners. Immunization agenda 2030: a global strategy to leave no one behind. Vaccine 2024 Apr 08;42 Suppl 1:S5-14 [FREE Full text] [doi: 10.1016/j.vaccine.2022.11.042] [Medline: 39004466]
5. Indonesia targets low vaccination areas to tackle decline in childhood immunization: during world immunization week, Indonesia recommits to increasing immunization. United Nations International Children's Emergency Fund. 2023 May 04. URL: https://www.unicef.org/indonesia/press-releases/indonesia-targets-low-vaccination-areas-tackle-decline-childhood-immunization [accessed 2023-08-05]
6. Impact of COVID-19 on routine immunization in Indonesia: quarterly assessment using Interactive Voice Recording (IVR) technology to assess the continuity of immunization services across all 34 provinces during the COVID-19 pandemic. United Nations International Children's Emergency Fund. URL: https://www.unicef.org/indonesia/reports/impact-covid-19-routine-immunization-indonesia [accessed 2023-09-10]
7. Petunjuk Teknis Bulan Imunisasi Anak Nasional (BIAN). Kemenkes. 2022 Jun 30. URL: https://ayosehat.kemkes.go.id/petunjuk-teknis-bulan-imunisasi-anak-nasional-bian [accessed 2023-08-05]
8. Routine immunization services during the COVID-19 pandemic. World Health Organization Regional Office for the Western Pacific. 2020 Apr 13. URL: https://apps.who.int/iris/bitstream/handle/10665/331925/Routine-immunization-services-COVID-19-eng.pdf [accessed 2023-09-10]
9. Yuliyanti S, Rifqi FM, Prasetyo RL, Nurhasanah E, Priambada AB, Abbas P. Analysis of basic immunization services during the COVID-19 pandemic at public health centers in Semarang. Open Access Maced J Med Sci 2021 Nov 26;9(E):1338-1344. [doi: 10.3889/oamjms.2021.7336]
10. Lim SS, Stein DB, Charrow A, Murray CJ. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. The Lancet 2008 Dec 13;372(9655):2031-2046. [doi: 10.1016/s0140-6736(08)61869-3]
11. Bosch-Capblanch X, Ronveaux O, Doyle V, Remedios V, Bchir A. Accuracy and quality of immunization information systems in forty-one low income countries. Trop Med Int Health 2009 Jan;14(1):2-10 [FREE Full text] [doi: 10.1111/j.1365-3156.2008.02181.x] [Medline: 19152556]
12. Ronveaux O, Rickert D, Hadler S, Groom H, Lloyd J, Bchir A, et al. The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems. Bull World Health Organ 2005 Jul;83(7):503-510 [FREE Full text] [Medline: 16175824]
13. Electronic immunization registry: practical considerations for planning, development, implementation and evaluation. Pan American Health Organization. 2017. URL: https://iris.paho.org/handle/10665.2/34865 [accessed 2023-08-11]

14. Dolan SB, Carnahan E, Shearer JC, Beylerian EN, Thompson J, Gilbert SS, et al. Redefining vaccination coverage and timeliness measures using electronic immunization registry data in low- and middle-income countries. Vaccine 2019 Mar 22;37(13):1859-1867 [FREE Full text] [doi: 10.1016/j.vaccine.2019.02.017] [Medline: 30808566]
15. Cetak biru strategi transformasi digital kesehatan 2024. Kementerian Kesehatan Republik Indonesia. 2021. URL: https://online.fliphtml5.com/egzdb/hsaa/#p=1 [accessed 2023-08-05]
16. Sehat Indonesiaku. Google Play. URL: https://play.google.com/store/apps/details?id=com.dto.sehatindonesiaku&hl=id [accessed 2023-08-04]
17. Kemenkes RI. Dashboard ASIK - Imunisasi Rutin - Saya sudah melakukan penginputan pada ASIK Mobile tapi capaian belum bertambah pada. Kemenkes. URL: https://faq.kemkes.go.id/faq/dashboard-asik-imunisasi-rutin-saya-sudah-melakukan-penginputan-pada-asik-mobile-tapi-capaian-belum-bertambah-pada-website [accessed 2025-02-01]
18. Danovaro-Holliday MC, Contreras MP, Pinto D, Molina-Aguilera IB, Miranda D, García O, et al. Assessing electronic immunization registries: the Pan American Health Organization experience. Rev Panam Salud Publica 2019;43:e28 [FREE Full text] [doi: 10.26633/RPSP.2019.28] [Medline: 31093252]
19. Rokom. Tak lagi dicatat manual, imunisasi anak akan terdata digital di Aplikasi Sehat IndonesiaKu (ASIK). Sehat Negeriku. 2022 May 18. URL: https://sehatnegeriku.kemkes.go.id/baca/umum/20220518/4039909/tak-lagi-dicatat-manual-imunisasi-anak-akan-terdata-digital-di-aplikasi-sehat-indonesiaku-asik/ [accessed 2023-08-06]
20. Rokom. Besok PeduliLindungi resmi bertransformasi menjadi SATUSEHAT mobile. Sehat Negeriku. 2023 Mar 10. URL: https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230228/2042474/besok-pedulilindungi-resmi-bertransformasi-menjadi-satusehat-mobile/ [accessed 2025-02-01]
21. Pedoman dan standar etik penelitian dan pengembangan kesehatan nasional. Komisi Etik Penelitian Dan Pengembangan Kesehatan Nasional Kementerian Kesehatan Republik Indonesia Tahun 2017. 2017 Nov 13. URL: https://kepk.poltekkestasikmalaya.ac.id/wp-content/uploads/2018/05/2017-KEPPKN-Standar-dan-Pedoman-.pdf [accessed 2023-09-10]
22. Capaian bulan imunisasi anak nasional tahap I dan tahap I tahun 2022. Kementerian Kesehatan Republik Indonesia. 2024. URL: https://drive.google.com/file/d/1RFE2cnJUbOBDsaiw0SSJAMEgG5YCVfhJ/view?usp=sharing [accessed 2025-02-23]
23. Bloland P, MacNeil A. Defining & assessing the quality, usability, and utilization of immunization data. BMC Public Health 2019 Apr 04;19(1):380 [FREE Full text] [doi: 10.1186/s12889-019-6709-1] [Medline: 30947703]
24. El-Khatib Z, El-Halabi S, Khdeir MA, Khader YS. Children immunization app (CImA), low-cost digital solution for supporting Syrian refugees in Zaatari camp in Jordan — general description. In: Proceedings of the IEEE/ACS 16th International Conference on Computer Systems and Applications. 2019 Presented at: AICCSA 2019; November 3-7, 2019; Abu Dhabi, United Arab Emirates. [doi: 10.1109/AICCSA47632.2019.9035211]
25. Cutts FT, Danovaro-Holliday MC, Rhoda DA. Challenges in measuring supplemental immunization activity coverage among measles zero-dose children. Vaccine 2021 Mar 01;39(9):1359-1363 [FREE Full text] [doi: 10.1016/j.vaccine.2020.11.050] [Medline: 33551302]
26. Wittenauer R, Dolan SB, Njoroge A, Onyango P, Owiso G, Rabinowitz P, et al. Usability and acceptability of electronic immunization registry data entry workflows from the health care worker perspective in Siaya, Kenya (part 3): pre-post study. JMIR Form Res 2023 Mar 30;7:e39383 [FREE Full text] [doi: 10.2196/39383] [Medline: 36995755]
27. Rolnick J. Aggregate health data in the United States: steps toward a public good. Health Informatics J 2013 Jun;19(2):137-151 [FREE Full text] [doi: 10.1177/1460458212462077] [Medline: 23715213]
28. Dang H, Dao S, Carnahan E, Kawakyu N, Duong H, Nguyen T, et al. Determinants of scale-up from a small pilot to a national electronic immunization registry in Vietnam: qualitative evaluation. J Med Internet Res 2020 Sep 22;22(9):e19923 [FREE Full text] [doi: 10.2196/19923] [Medline: 32960184]
29. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Jun 19;6:54. [doi: 10.1186/s40537-019-0217-0]
30. Stockwell MS, Natarajan K, Ramakrishnan R, Holleran S, Forney K, Aponte A, et al. Immunization data exchange with electronic health records. Pediatrics 2016 Jun;137(6):e20154335. [doi: 10.1542/peds.2015-4335] [Medline: 27244856]
31. McGreevy S, Murray M, Montero L, Gibson C, Comfort B, Barry M, et al. Assessing the immunization information system and electronic health record interface accuracy for COVID-19 vaccinations. JAMIA Open 2023 Apr 13;6(2):ooad026 [FREE Full text] [doi: 10.1093/jamiaopen/ooad026] [Medline: 37063406]
32. Groom HC, Crane B, Naleway AL, Weintraub E, Daley MF, Wain K, et al. Monitoring vaccine safety using the vaccine safety datalink: assessing capacity to integrate data from immunization information systems. Vaccine 2022 Jan 31;40(5):752-756 [FREE Full text] [doi: 10.1016/j.vaccine.2021.12.048] [Medline: 34980508]
33. Accelerating the app economy in Indonesia: Android and Google Play's impact in Indonesia. Access Partnership. 2023 Aug. URL: https://accesspartnership.com/wp-content/uploads/2023/08/Accelerating-the-app-economy-in-indonesia.pdf [accessed 2023-09-13]
34. Statistik telekomunikasi Indonesia 2021. Badan Pusat Statistik Indonesia. 2022 Sep 7. URL: https://www.bps.go.id/publication/2022/09/07/bcc820e694c537ed3ec131b9/statistik-telekomunikasi-indonesia-2021.html [accessed 2023-08-13]
35. Smartphone users in Indonesia. Start.io. URL: https://www.start.io/audience/smartphone-users-in-indonesia [accessed 2023-08-13]

36. Hasan S, Yousuf MM, Farooq M, Marwah N, Andrabi SA, Kumar H. e-vaccine: an immunization app. In: Proceedings of the 2nd International Conference on Intelligent Engineering and Management. 2021 Presented at: ICIEM 2021; April 28-30, 2021; London, UK. [doi: 10.1109/iciem51511.2021.9445386]
37. Sullivan E, Masood T, Javed W, Bagshaw K, Ollis S, Regmi P, et al. Electronic immunization information systems: a case report of lessons learned from implementation in Pakistan. Mhealth 2020 Jul 05;6:31 [FREE Full text] [doi: 10.21037/mhealth.2020.01.07] [Medline: 32632369]
38. Wilson K, Atkinson KM, Deeks SL, Crowcroft NS. Improving vaccine registries through mobile technologies: a vision for mobile enhanced Immunization information systems. J Am Med Inform Assoc 2016 Jan;23(1):207-211 [FREE Full text] [doi: 10.1093/jamia/ocv055] [Medline: 26078414]
39. Siddiqi DA, Abdullah S, Dharma VK, Shah MT, Akhter MA, Habib A, et al. Using a low-cost, real-time electronic immunization registry in Pakistan to demonstrate utility of data for immunization programs and evidence-based decision making to achieve SDG-3: insights from analysis of Big Data on vaccines. Int J Med Inform 2021 May;149:104413. [doi: 10.1016/j.ijmedinf.2021.104413] [Medline: 33652259]
40. Seymour D, Werner L, Mwansa FD, Bulula N, Mwanyika H, Dube M, et al. Electronic immunization registries in Tanzania and Zambia: shaping a minimum viable product for scaled solutions. Front Public Health 2019 Aug 7;7:218 [FREE Full text] [doi: 10.3389/fpubh.2019.00218] [Medline: 31440494]
41. Pratama MA, Cahyadi AT. Effect of user interface and user experience on application sales. IOP Conf Ser Mater Sci Eng 2020;879:012133. [doi: 10.1088/1757-899X/879/1/012133]
42. Carnahan E, Ferriss E, Beylerian E, Mwansa FD, Bulula N, Lyimo D, et al. Determinants of facility-level use of electronic immunization registries in Tanzania and Zambia: an observational analysis. Glob Health Sci Pract 2020 Sep;8(3):488-504. [doi: 10.9745/ghsp-d-20-00134]
43. Shen Z, Xue C, Wang H. Effects of users' familiarity with the objects depicted in icons on the cognitive performance of icon identification. Iperception 2018 Jun 11;9(3):2041669518780807 [FREE Full text] [doi: 10.1177/2041669518780807] [Medline: 29977490]
44. Nicol E, Bradshaw D, Phillips T, Dudley L. Human factors affecting the quality of routinely collected data in South Africa. Stud Health Technol Inform 2013;192:788-792. [Medline: 23920665]
45. Dharmayuda AA, Wulandari LP. Analisis beban kerja dokter umum di Puskesmas kota Denpasar dengan menggunakan metode workload indicators of staffing need. Public Health Prevent Med Arch 2015 Jul;3(1):16 [FREE Full text] [doi: 10.15562/phpma.v3i1.95]
46. Hlaing T, Myint ZM. Factors affecting data quality of health management information system at township level, Bago region, Myanmar. Int J Community Med Public Health 2022 Feb 28;9(3):1298-1307. [doi: 10.18203/2394-6040.ijcmph20220686]
47. Purwanti E, Ediyono S. The relationship between heavy workload and the nurse's performance. Int J Med Sci Clin Res Stud 2022 Jun 01;2(6):445-450. [doi: 10.47191/ijmscrs/v2-i6-01]
48. Glèlè Ahanhanzo Y, Ouedraogo LT, Kpozèhouen A, Coppieters Y, Makoutodé M, Wilmet-Dramaix M. Factors associated with data quality in the routine health information system of Benin. Arch Public Health 2014 Jul 28;72(1):25 [FREE Full text] [doi: 10.1186/2049-3258-72-25] [Medline: 25114792]
49. Khobi JA, Mtebe JS, Mbelwa JT. Factors influencing district health information system usage in Sierra Leone: a study using the technology-organization-environment framework. Electron J Inf Syst Dev Ctries 2020 Nov;86:e12140. [doi: 10.1002/isd2.12140]
50. Terbaru! Ketentuan pelaksanaan vaksin booster (Per 27 Januari 2022). Satuan Tugas Penanganan COVID-19. 2022 Jan 29. URL: https://covid19.go.id/tenaga-kesehatan/terbaru-ketentuan-pelaksanaan-vaksin-booster-27-januari-2022 [accessed 2023-08-14]
51. Nugraha RR, Miranda AV, Ahmadi A, Lucero-Prisno DE. Accelerating Indonesian COVID-19 vaccination rollout: a critical task amid the second wave. Trop Med Health 2021 Sep 22;49(1):76. [doi: 10.1186/s41182-021-00367-3] [Medline: 34548099]
52. COVID-19 vaccine post introduction evaluation (cPIE): Indonesia country report. World Health Organization. 2022. URL: https://cdn.who.int/media/docs/default-source/searo/indonesia/indonesia-cpie-report.pdf [accessed 2023-08-14]
53. Classification of digital health interventions v 1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2023-08-08]
54. Namageyo-Funa A, Samuel A, Bloland P, Macneil A. Considerations for the development and implementation of electronic immunization registries in Africa. Pan Afr Med J 2018;30:81. [doi: 10.11604/pamj.2018.30.81.11951]
55. Soemartono T. The dynamic of e-KTP evaluation program in DKI Jakarta. Int J Adm Sci Organ 2014;20(2):90-97. [doi: 10.20476/jbb.v20i2.3070]
56. Tangkilisan W, Sjam JM, Supriadi T, Marota R. Legal problems of the misuse of electronic identification card (E-Ktp) data in the Indonesian general election. J Posit Sch Psychol 2022;6(6):5851-5866.

57. Beyond unicorns: harnessing digital technologies for inclusion in Indonesia. World Bank Group. 2021 Jul 28. URL: https://www.worldbank.org/en/country/indonesia/publication/beyond-unicorns-harnessing-digital-technologies-for-inclusion-in-indonesia [accessed 2023-08-14]
58. Telehealth and digital inclusion in Indonesia. High-Quality Technical Assistance for Results. 2022. URL: https://archive.ids.ac.uk/heart/wp-content/uploads/2022/01/Telehealth-and-Digital-Inclusion-in-Indonesia-Executive-Summary.pdf [accessed 2023-08-14]
59. Mitgang EA, Blaya JA, Chopra M. Digital health in response to COVID-19 in low- and middle-income countries: opportunities and challenges. Glob Policy 2021 Jul;12(Suppl 6):107-109 [FREE Full text] [doi: 10.1111/1758-5899.12880] [Medline: 34230840]
60. Akhtar MD, Mathur G, Kravchenko O, Rakhra M. Connecting the unconnected: bridging the digital divide with affordable satellite through enabled smartphones. E3S Web Conf 2023 Nov 30;453:01054. [doi: 10.1051/e3sconf/202345301054]
61. Aisyah DN, Setiawan AH, Lokopessy AF, Faradiba N, Setiaji S, Manikam L, et al. The information and communication technology maturity assessment at primary health care services across 9 provinces in Indonesia: evaluation study. JMIR Med Inform 2024 Jul 18;12:e55959 [FREE Full text] [doi: 10.2196/55959] [Medline: 39037345]

Abbreviations

ASIK: Aplikasi Sehat IndonesiaKu
BIAN: Bulan Imunisasi Anak Nasional
DPT-HB-Hib: diphtheria, pertussis, tetanus, hepatitis B, and Haemophilus influenzae type B
EIR: electronic immunization registry
IPV: inactivated polio vaccine
MoH: Ministry of Health
MR: measles-rubella
OPV: oral polio vaccine
SMILE: Electronic Logistic Management and Information System
UI: user interface
UX: user experience

Edited by T de Azevedo Cardoso; submitted 21.10.23; peer-reviewed by P Mechael, A Rusibamayila, C Weerth; comments to author 08.04.24; revised version received 01.07.24; accepted 19.12.24; published 27.03.25.

Please cite as:
Aisyah DN, Utami A, Rahman FM, Adriani NH, Fitransyah F, Endryantoro MTA, Hutapea PY, Tandy G, Manikam L, Kozlakidis Z
Using an Electronic Immunization Registry (Aplikasi Sehat IndonesiaKu) in Indonesia: Cross-Sectional Study
Interact J Med Res 2025;14:e53849
URL: https://www.i-jmr.org/2025/1/e53849
doi:10.2196/53849
PMID:

How to Design Electronic Case Report Form (eCRF) Questions to Maximize Semantic Interoperability in Clinical Research

Eugenia Rinaldi*, MSc; Caroline Stellmach*, MSc; Sylvia Thun, MD

Core Unit Digital Medicine and Interoperability, Berlin Institute of Health, Charité - Universitätsmedizin Berlin, Luisenstr. 65, Berlin, Germany
*these authors contributed equally

Corresponding Author:
Eugenia Rinaldi, MSc
Core Unit Digital Medicine and Interoperability, Berlin Institute of Health, Charité - Universitätsmedizin Berlin, Luisenstr. 65, Berlin, Germany

Abstract

Case report forms (CRFs) are the instruments used by research organizations worldwide to collect information about patients and study participants with the purpose of answering specific questions, assessing the efficacy and safety of medical products, and in general improving prevention and treatment in health care. To obtain significant research results out of the collected data, CRFs should be designed following the recommendations issued by regulatory authorities. However, we believe that semantic interoperability in CRFs has not yet been properly addressed. Within an international consortium comprising several COVID-19 cohorts, we scrutinized the questions included in the different CRFs with the purpose of establishing semantic interoperability across the different study data elements so that data could be merged and jointly analyzed. We realized that similar concepts were structured very differently across the different CRFs, making it hard to find and match the information. Based on the experience acquired, we developed 5 guiding principles on how to design CRFs to support semantic interoperability and increase data quality while also facilitating the sharing of data. Our aim in this viewpoint is to provide general suggestions that, in our opinion, should support researchers in designing CRFs. We conclude by urging authorities to establish an international coordination board for standards and interoperable clinical study data with competence in clinical data, interoperability standards, and data protection as part of a preparedness plan for future pandemics or other health threats.

(Interact J Med Res 2025;14:e51598) doi:10.2196/51598

KEYWORDS

case report form; CRF; interoperability; standard data model; data format; metadata; core data elements; data quality

Introduction

Since the onset of the COVID-19 pandemic, we have witnessed the emergence of numerous studies worldwide aimed at deepening our understanding of SARS-CoV-2 infection and enhancing treatment strategies [1]. The urgency to produce meaningful research results in a relatively short timeframe has underscored the importance of efficiently merging data from diverse studies. However, differences in languages, formats, and terminologies often complicate the sharing and integration of data.

The European project ORCHESTRA [2], which sought to build a pan-European cohort of COVID-19 patients, confronted the significant challenge of harmonizing data from various cross-country studies [3]. Achieving interoperability among disparate datasets is critical not only for supporting ongoing research but also for bolstering preparedness against future global health crises. This aligns with the European Commission's objectives under the European Health Data Space regulation [4], which emphasizes improving data accessibility and integration across member states.

Case report forms (CRFs) are essential tools employed by research organizations worldwide to gather detailed information from patients and study participants. These forms are designed to address specific research questions, evaluate the efficacy and safety of medical interventions, and ultimately advance prevention and treatment in health care.

A clinical study is designed to answer one or several research questions based on the analysis of data collected from patients that have been enrolled and are being observed or are partaking in specific interventions following the study protocol. CRFs, either in paper form or electronic format, are the instruments used to collect study data and form the basis of any subsequent statistical analysis. Electronic CRFs (eCRFs) are preferred over paper CRFs because data can automatically be stored in a digital format and immediately used, removing inaccuracies derived from the interpretation and transcription of handwriting. Additionally, if properly structured, the digital format intrinsically offers great potential for data objects to be more findable, accessible, interoperable, and reusable, in other words, more FAIR [5], than paper. The design of eCRFs is crucial for the outcome of a study [6]. Therefore, it should be optimized to enhance data quality and data interoperability.

Ideally, a standard operating procedure is established initially for designing eCRFs. Common recommendations for eCRF design include suggestions to reduce data entry errors and ambiguity in the interpretation of variables such as maximizing the use of coded questions and answer lists and minimizing the use of free text answers; using built-in consistency checks for admissible ranges and plausible date checks; facilitating data entry using branching logic strategies; specifying units of measurement (particularly for laboratory parameters, but also for vital signs, etc); adopting standard data formats; using (and reusing) published Common Data Elements, if available, and unambiguous temporal reference (eg, before or during infection).

Regulatory authorities and international expert organizations, such as the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use and the Society for Clinical Data Management, have published detailed guidance on how to design CRFs, placing their focus on ensuring accuracy, utility, and poignancy of layout and content[7],[8] , [9]. However, the idea of designing CRF variables to facilitate standardization and reusability in the context of data interoperability has remained unaddressed by these expert institutions.

In the simplest setup, a study will be limited to 1 cohort, enrolling patients within one country, across several participating medical offices or hospitals. In more complex cases, patients are enrolled across several cohorts located in multiple countries. Regardless of the complexity of the study, interoperability and quality of data should always be considered high-priority objectives when designing the collection forms. By so doing, study results are more reliable, and collected data are ready for potential secondary use.

The COVID-19 pandemic, along with disease outbreaks caused by other viruses such as human monkeypox, Zika, and Ebola, has highlighted the need for international collaboration in terms of research. They have led to large-scale clinical studies conducted in the private sector and by public research consortia [10]. Multi-country and multi-cohort retrospective studies come with more challenges, as they generally need to combine data that were collected in different formats and, at times, different languages. This means that even when variables cover similar information, extensive transformation or translation activities are required before merging data can take place. This highlights the need for coherent study protocols across research groups and across countries, based on common formats and terminologies. Here is where data standardization and harmonization are the key to enabling quality data that can be merged easily without resource-heavy transformation activities, thus expediting analysis and gaining timely insights [11,12].

Aim

Our aim in this viewpoint is to provide researchers with general suggestions in designing CRFs that, in our opinion, should support interoperability, reduce ambiguity, improve data quality, and facilitate data exchange across different systems.

The ORCHESTRA Project

ORCHESTRA was funded by the European Commission during the COVID-19 pandemic and ran until the end of 2024. It aimed at creating a pan-European cohort of COVID-19 patients to study the disease, the efficacy of the treatment, and the long-term effects on general and fragile population as well as on health care workers. We were responsible for establishing interoperability within the European ORCHESTRA project. Partners in ORCHESTRA followed the goal to merge data from different clinical studies to generate new knowledge about multiple aspects of COVID-19 [3]. We examined over 3700 variables (comprised of questions and answers) with the objective of identifying similar information across studies that could be matched and analyzed jointly. This way, we compared the variety of approaches used across 7 different COVID-19 studies to investigate similar health care concepts. Within our project task, we associated international identifying codes from the most pertinent standard health terminologies to the questions and answers included in 7 clinical studies in ORCHESTRA. Ambiguous or complex wording found in the CRF variables needed to be evaluated as part of this mapping activity as well.

Based on our experience of being involved in large-scale national and international research projects, we believe that apart from ensuring accuracy and quality of collected data, CRF design should also maximize semantic interoperability. That way, time- and cost-efficient merging, analysis, and sharing of data can be facilitated. Our conviction is supported by the report published by the Joint Action Towards the European Health Data Space, which is a European initiative that developed principles for the secondary use of health data [13] and that places semantic interoperability as one of the operational objectives to achieve excellence of data quality.

To address the pressing need for streamlined data exchange and integration in clinical research, we have formulated 5 guiding principles that should be considered when designing CRFs. The principles address the need to harmonize data, unambiguously identify variables, associate clinical concepts with international identifiers, promote data quality, and enhance semantic coherence.

We believe that the application of the proposed principles would enhance semantic interoperability and support the exchange of information across different research groups.

Five Guiding Principles to Enhance Interoperability of CRFs

Following our aim statement, we propose 5 guiding concepts aimed at increasing semantic interoperability of CRF data and quality of collected information.

Harmonize Data

Concept 1: Creation or Reuse of Core Data Elements (CDEs)

Recurring information across clinical trials that are general in nature or specific to a specific disease should be identified and shared in a common format across the scientific community.

Data concerning demographics or clinical evaluation of patients, for example, is collected in many studies and should ideally be standardized to create a uniform format. Frequently collected disease progression and outcome information should also be identified and grouped into well-defined disease-specific core data elements (CDEs). CDEs can then be published for re-use by researchers worldwide.

Granted, the National Institutes of Health (NIH) have been focusing on developing CDEs for over 20 years. Yet, the adoption of CDEs is facing challenges. Reasons for low adoption by research projects, among others, are established data collection practices and ambiguous interpretations and implementations of health care concepts and CDEs.

The latter difficulty can, however be overcome by associating a common standard terminology to data elements to remove any ambiguity of meaning.

The increasing need to combine data in order to address global threats to life will eventually have to gain greater weight than maintaining localized, established practices.

National and international harmonization efforts such as the International Patient Summary (IPS) [14], the European Electronic Health Record Exchange Format [15], or the United States Core Data for Interoperability (USCDI) [16] should be considered to maximize reusability of data.

Unambiguously Identify Variables

Concept 2: Use of Standards in CRF Metadata

Following the FAIR principles, metadata is of foremost importance for the quality re-use of information. Metadata should ideally include references to international terminology codes that unambiguously represent each concept. Whenever possible, all CRF variables should be handled as close-ended questions. In case of measurement or observations, Logical Observation Identifiers Names and Codes (LOINC) should be used. LOINC offers the possibility of defining data elements without ambiguity by specifying with one code all the details relative to the observation to be performed.

When no LOINC codes are available, other international terminologies should be considered, such as the National Cancer Institute's Thesaurus for genomics data.

Qualitative answers should be restricted to defined (coded) value sets and identified with the appropriate terminology codes, such as those provided by Systematized Medical Nomenclature for Medicine–Clinical Terminology (SNOMED CT) [17]. The Anatomical Therapeutic Chemical Classification System can be used to describe drugs and chemicals, and the *International Classification of Diseases* to report diseases and disorders.

CRFs often include more complex questions that cover several informational components. That can include details about time, location, situation, etc. Hence, mapping several concepts to one semantic code can at times be difficult. This is a challenge that the postcoordination expression in SNOMED CT may help to solve in some contexts. However, integrating postcoordination into CRFs appears less feasible [18].

Terminology bindings proposed by national or international harmonization initiatives such as those mentioned in Concept 1 should be reflected in the metadata of the study data elements whenever possible. It should also be noted that the metadata could also include information on the format of the data. That can be achieved by mapping study elements to standards such as the Health Level Seven's (HL7) Fast Healthcare Interoperability Resources (FHIR) standard or to the Observational Medical Outcomes Partnership Common Data Model.

The combined use of semantic and syntactic standards would further support interoperability.

Associate Clinical Concepts With International Identifiers

Concept 3: One Concept at a Time (Unless It Is Part of a Questionnaire or Index)

The progressive adoption of FHIR by initiatives aiming to harmonize health data, such as IPS and USCDI, would suggest that information should follow the modular structure and be as precise as possible.

Hence, our second recommendation would be to only include one concept in a question. We have, for example, seen the following question in a COVID-19–related study CRF: asking enrolled patients whether they have had "Changes in or a loss of smell." In this case, we propose splitting the question into 2 variables: "Changes in smell" and "Loss of smell." With this approach, each variable could be represented by a specific semantic code. Additionally, splitting the question into 2 would also facilitate accurate analysis.

We acknowledge that this accuracy has to be balanced against the manageability of the length of a CRF. It should not lead to the creation of overly long CRFs, but rather to a focus on accuracy and key variables needed for analysis.

However, it is also important to note that at times some concepts which are included in questions could be removed without lowering accuracy of the question. For example, information like time points or target patients could be put as a header or as instructions within the CRF.

Of course, the ultimate decision on what constitutes the most relevant concepts to be included in the CRF questions always lies with the scientific group or the principal investigator or the sponsor.

Promote Data Quality

Concept 4: Accurate Wording

Initiatives such as IPS and the USCDI provide guidance on the use of patient-related core data elements. However, often CRFs

require the use of more study-specific variables. The wording of variables (questions and answers) should be carefully phrased to provide all and only the necessary information so that study nurses and respondents understand exactly what is asked. Precisely worded questions (and answer options) will increase the quality of data. This is to ensure that the data collected in response to these variables will be comparable and ready to be merged. In the case of laboratory examinations, the use of LOINC codes can be very useful because it automatically includes all the information necessary to remove ambiguity, that is, the methodology used, specimen type, or whether the expected result is qualitative or quantitative, unit information, etc.

If language translations of CRF variables are required, these should be meticulously performed. Additionally, a quality check on the wording and meaning of the translated variables is recommended and should be implemented, ideally by native speakers.

Enhance Semantic Coherence

Concept 5: Answer Options Should All Be Semantically Coherent, and Units of Measures Clearly Stated and Identified With the Unified Code for Units of Measure Units

Another aspect of standardization relates to the answer options that are the second component of a CRF variable, after the question. It is important to maintain coherent semantic coding of answers as well. We recommend that, during the design of CRFs, semantic codes be used for mapping answers instead of assigning generic numeric identifiers that lack specificity. This would provide an unambiguous and reusable representation of the answer concept. Furthermore, coding would help the precision of the information and the quality of data by highlighting inconsistencies such as value set options not being semantically aligned.

For example, the question “What kind of swab test was performed?” should not include “throat” and “PCR” as answers in the same value set. In this case, the use of a terminology system like SNOMED CT clearly shows that the codes of the 2 concepts belong to different semantic categories (“body structure” and “procedure” respectively). Therefore, the question might be equivocal and could lead to unclear results.

In addition, in case of variables that describe quantitative (laboratory) measurements, such as “Dose of immunosuppressive medication taken per day” “Body weight” or “Glucose concentration measured” units of measure should be clearly stated and identified with Unified Code for Units of Measure codes [19,20].

Summary and Considerations

Based on our experience in standardizing and harmonizing CRF variables from different protocols on COVID-19, we have presented five concepts aimed at improving CRF design and enhancing interoperability of clinical study data: (1) the creation of or (2) use of already existing standardized data objects can save time and help establish alignment and comparability with other research datasets, and (3) data quality can significantly be improved by paying attention to the fact that each CRF question and every answer option only contains 1 concept, (4) that variables are accurately phrased, and (5) that answer options are coherent, or in case of numerical results, that clearly defined units are included.

Probably, guidelines for reporting multiple concept codes in the metadata should be established, as this is a common occurrence in clinical research. In many cases, the possibility to code complex questions as coded questionnaires is very helpful. We believe it would be important to collaboratively address this difficulty of coding complex variables for the clinical study context.

The problem of not using interoperability standards is invisible to many researchers. That is because often, the advantages of such use become evident only when the need to merge data arises. Ideally, standards should be introduced already during the design phase of CRFs. Unfortunately, since the importance of data standards is still not adequately known, their implementation might be seen by some as a hassle that slows down or even limits the development of CRFs. In our opinion, a cultural change based on education and information in this field is needed.

It is necessary to abandon the idea of doing research in silos with data collected in incompatible formats by different research groups. Common data elements and their format should be identified, agreed on, and promoted by relevant national and international authorities. On the other hand, a scenario where clinicians are unwillingly responsible for standardization should be avoided, considering the already existing strain on their time and understandable gaps in expertise. New roles in health care are needed with expertise in digital medicine to enable interoperability of data and facilitate their integration within a wider eHealth ecosystem where data are being collected from different digital solutions and in a cross-country context.

The use of a common exchange format in combination with standard terminologies for study data elements in an eCRF would complete the interoperable data model of clinical research information. The innovative standard that is increasingly being adopted in the health care environment is HL7 FHIR. Thanks to FHIR’s innovative modular organization of information, a particularly efficient exchange of data is enabled. Its adoption in the clinical research environment is still low, but we expect this to change in the future. That is because the need to streamline activities and integrate information from electronic patient records or other medical devices into clinical research is progressively becoming more evident [21,22]. In fact, a dedicated HL7 working group is focusing on the design of FHIR resources to conduct clinical research more effectively [23].

In this context, as mentioned before, the European Commission has proposed a regulation for the European Health Data Space to support the interoperability of data in healthcare and in research. Forthcoming implementing acts will provide specifications for the exchange format of data to support cross-system and cross-border portability. In the United States, the Office of the National Coordinator for Health Information Technology promotes the USCDI [16], a set of health data

elements divided into thematic classes to support information exchange. In general, collaborations to improve interoperability are fortunately increasing, including, for example, the Observational Health Data Sciences and Informatics (OHDSI) and the NIH concerning the maternal health data [24] or the employment of the Observational Medical Outcomes Partnership Common Data Model by the NIH research program "All of Us" [25]. Very important is also the OHDSI and European Medicines Agency collaboration in the project DARWIN [26], and the HL7 Vulcan project bringing together OHDSI and HL7 FHIR [27].

In conclusion, the COVID-19 pandemic has revealed to the broad community how important it is to quickly analyze large amounts of data, develop vaccines, and assess their safety and efficacy. We therefore need to facilitate the exchange of information in the context of global health challenges (including cancer, infectious disease, and rare diseases) and implement standardization of clinical study data collection. Additionally, the establishment of an international coordination board for standards and interoperable clinical study data with competence in clinical data, interoperability standards, and data protection should be part of a preparedness plan to face future pandemics and other health threats. This proposed coordination board, in coordination with ongoing international initiatives, could be instituted at the regional level and associated with large funding bodies and policy makers (ie, European Commission within the European Union and NIH in the United States) or at the pan-regional level, for example, as part of the World Health Organization.

Acknowledgments

We would like to thank Prof Evelina Tacconelli for her invaluable support. The guiding principles were conceived based on the activities performed within the ORCHESTRA project, which has received funding from the European Union's Horizon 2020 research and innovation program (grant 101016167).

Data Availability

The metadata definitions on which we based our comment are publicly available on the standard-enabling platform ART-DECOR [28].

Authors' Contributions

ER and CS contributed equally conceiving and designing the guiding principles. ST has revised the paper critically and made important contributions.

Conflicts of Interest

None declared.

References

1. Sass J, Bartschke A, Lehne M, et al. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. Research Square. Preprint posted online on Dec 21, 2020. [doi: 10.21203/rs.3.rs-51348/v3]
2. ORCHESTRA Cohort. URL: https://orchestra-cohort.eu/ [accessed 2024-12-08]
3. Rinaldi E, Stellmach C, Rajkumar NMR, et al. Harmonization and standardization of data for a pan-European cohort on SARS- CoV-2 pandemic. NPJ Digit Med 2022 Jun 14;5(1):75. [doi: 10.1038/s41746-022-00620-x] [Medline: 35701537]
4. Kessissoglou IA, Cosgrove SM, Abboud LA, Bogaert P, Peolsson M, Calleja N. Are EU member states ready for the European Health Data Space? Lessons learnt on the secondary use of health data from the TEHDAS Joint Action. Eur J Public Health 2024 Dec 1;34(6):1102-1108. [doi: 10.1093/eurpub/ckae160] [Medline: 39514641]
5. Solle D. Be FAIR to your data. Anal Bioanal Chem 2020 Jul;412(17):3961-3965. [doi: 10.1007/s00216-020-02526-7] [Medline: 32300841]
6. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspect Clin Res 2014 Oct;5(4):159-166. [doi: 10.4103/2229-3485.140555] [Medline: 25276625]
7. Guideline for good clinical practice E6(R2). European Medicines Agency. 2016 Jan 12. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-good-clinical-practice-e6r2-step-5_en.pdf [accessed 2023-02-14]
8. Good clinical data management practices. Society for Clinical Data Management. 2023 Oct. URL: https://scdm.org/wp-content/uploads/2019/10/21117-Full-GCDMP-Oct-2013.pdf [accessed 2023-03-20]
9. Richesson RL, Nadkarni P. Data standards for clinical research data collection forms: current status and challenges. J Am Med Inform Assoc 2011 May 1;18(3):341-346. [doi: 10.1136/amiajnl-2011-000107] [Medline: 21486890]
10. Beall RF, Moradpour J, Hollis A. The private versus public contribution to the biomedical literature during the COVID-19, Ebola, H1N1, and Zika public health emergencies. PLoS One 2021;16(10):e0258013. [doi: 10.1371/journal.pone.0258013] [Medline: 34679120]

11. Lin CH, Wu NY, Liou DM. A multi-technique approach to bridge electronic case report form design and data standard adoption. J Biomed Inform 2015 Feb;53:49-57. [doi: 10.1016/j.jbi.2014.08.013] [Medline: 25200473]
12. Fegan G, Cheah PY, Data Sharing Working Group. Electronic address: WorkingGroups@covid19crc.org. Solutions to COVID-19 data sharing. Lancet Digit Health 2021 Jan;3(1):e6. [doi: 10.1016/S2589-7500(20)30273-9] [Medline: 33735070]
13. TEHDAS. URL: https://tehdas.eu/ [accessed 2025-02-24]
14. Kay S, Cangioli G, Nusbaum M. The international patient summary standard and the extensibility requirement. Stud Health Technol Inform 2020 Sep 4;273:54-62. [doi: 10.3233/SHTI200615] [Medline: 33087592]
15. COMMISSION RECOMMENDATION (EU) 2019/ 243 - of 6 February 2019 - on a European Electronic Health Record exchange format. : Official Journal of the European Union; 2019 URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019H0243#:~:text=(1)%20This%20Recommendation%20sets%20out,health%20data%20in%20the%20Union [accessed 2025-02-24]
16. United States Core Data for Interoperability (USCDI). Office of the National Coordinator for Health Information Technology. URL: http://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi [accessed 2024-05-15]
17. Bietenbeck A, Streichert T. Preparing laboratories for interconnected health care. Diagnostics (Basel) 2021 Aug 17;11(8):1487. [doi: 10.3390/diagnostics11081487] [Medline: 34441421]
18. Cornet R, Nyström M, Karlsson D. User-directed coordination in SNOMED CT. Stud Health Technol Inform 2013;192:72-76. [Medline: 23920518]
19. Unified Code for Units of Measure (UCUM). National Library of Medicine. URL: https://ucum.nlm.nih.gov/ [accessed 2023-03-20]
20. Rajput AM, Ballout S, Drenkhahn C. Standardizing the unit of measurements in LOINC-coded laboratory tests can significantly improve semantic interoperability. Stud Health Technol Inform 2020 Nov 23;275:234-235. [doi: 10.3233/SHTI200733] [Medline: 33227779]
21. Pétavy F, et al. Global standardization of clinical research data. Appl Clin Trials 2019;28(4):2023 [FREE Full text]
22. Kush RD, Nordo AH. In: Richesson RL, Andrews JE, editors. Data Sharing and Reuse of Health Data for Research: Springer International Publishing; 2019:379-401. [doi: 10.1007/978-3-319-98779-8_18]
23. Leroux H, Denney CK, Hastak S, Glover H. A framework for representing clinical research in FHIR. Presented at: 12th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences; Dec 10-11, 2019; Edinburgh, Scotland URL: https://ceur-ws.org/Vol-2849/paper-04.pdf [accessed 2025-02-24]
24. Johns Hopkins awarded NIH maternal health research data innovation and coordination hub. Biomedical Informatics and Data Science (BIDS). URL: http://dhsi.med.jhmi.edu/news/johns-hopkins-awarded-nih-maternal-health-research-data-innovation-and-coordination-hub [accessed 2024-05-16]
25. Ronquillo JG, Lester WT. Precision medicine landscape of genomic testing for patients with cancer in the National Institutes of Health All of Us Database Using Informatics Approaches. JCO Clin Cancer Inform 2022 Jan;6:e2100152. [doi: 10.1200/CCI.21.00152] [Medline: 34985965]
26. Dernie F, Corby G, Robinson A, et al. Standardised and reproducible phenotyping using distributed analytics and tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). Pharmacoepidemiol Drug Saf 2024. [doi: 10.1002/pds.70042]
27. FHIR to OMOP FHIR IG. URL: https://build.fhir.org/ig/HL7/fhir-omop-ig/ [accessed 2024-05-16]
28. Orchestra - project information. ART-DECOR. URL: https://art-decor.org/art-decor/decor-project--orch [accessed 2025-02-24]

Abbreviations

CDE: core data element
CRF: case report form
eCRF: electronic case report form
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven
IPS: International Patient Summary
NIH: National Institutes of Health
OHDSI: Observational Health Data Sciences and Informatics
SNOMED CT: Systematized Medical Nomenclature for Medicine–Clinical Terminology
USCDI: United States Core Data for Interoperability

Edited by TDA Cardoso; submitted 04.08.23; peer-reviewed by B Senst, B Feng, C Asuzu, F Amar; revised version received 08.12.24; accepted 08.01.25; published 03.03.25.

Please cite as:
Rinaldi E, Stellmach C, Thun S
How to Design Electronic Case Report Form (eCRF) Questions to Maximize Semantic Interoperability in Clinical Research
Interact J Med Res 2025;14:e51598
URL: https://www.i-jmr.org/2025/1/e51598
doi:10.2196/51598

Viewpoint

The Evolution of Uroflowmetry and Bladder Diary and the Emerging Trend of Using Home Devices From Hospital to Home

Ming-wei Li[1]; Yao-Chou Tsai[1,2]; Stephen Shei-Dei Yang[1,2]; Yuan-Hung Pong[3]; Yu-Ting Tsai[4]; Vincent Fang-Sheng Tsai[1,5]

[1]Division of Urology, Department of Surgery, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan

[2]School of Medicine, Tzu Chi University, Hualien, Taiwan

[3]Department of Urology, Ten Chen Hospital, Taoyuan City, Taiwan

[4]PhD Program of Mechanical and Aeronautical Engineering, Feng Chia University, Taichung, Taiwan

[5]Department of Urology, National Taiwan University Hospital, Taipei city, Taiwan

Corresponding Author:
Vincent Fang-Sheng Tsai
Division of Urology, Department of Surgery
Taipei Tzu Chi Hospital
Buddhist Tzu Chi Medical Foundation
16F, No 289, Jianguo Road
New Taipei City, 231016
Taiwan
Phone: 886 952 691807
Email: ntubala@gmail.com

Abstract

Although uroflowmetry and bladder diaries are widely used for noninvasive evaluation of lower urinary tract symptoms, they still have limitations in diagnostic capability and users' convenience. The aim of this paper is to discuss potential solutions by reviewing (1) the evolution and current clinical use of uroflowmetry and bladder diary, including clinical guidelines, daily practice applications, and their historical development; (2) a growing trend toward using home devices with various technologies; and (3) a comprehensive comparison of the strengths and weaknesses of these home devices. In our opinion, the following points can be highlighted: (1) the emerging trend of using home devices can enhance diagnostic capabilities through repeated measurements and the convenience of at-home testing and (2) home devices, which provide both frequency-volume and uroflowmetry information, have the potential to transform the management of lower urinary tract symptoms.

(Interact J Med Res 2025;14:e66694) doi:10.2196/66694

KEYWORDS

lower urinary tract symptoms; uroflowmetry; bladder diary; home devices; bladder; noninvasive; evaluations; viewpoint; diagnostic; mobile health

Introduction

As the population ages, there has been an increase in patients reporting lower urinary tract symptoms (LUTS) in recent years. To objectively assess the function of the lower urinary tract, uroflowmetry and bladder diaries (BD) are commonly used noninvasive examinations for those experiencing LUTS.

Uroflowmetry can measure various parameters during the voiding phase, including maximum flow rate (Qmax), and voided volume, which are essential metrics. Additional parameters, such as flow pattern, time to Qmax, flow time, and average flow rate, provide further insights [1]. This makes uroflowmetry a comprehensive and objective tool for evaluating the voiding phase. Consequently, uroflowmetry is recommended in the guidelines of both the American Urological Association and the European Association of Urology for assessing LUTS [2-4].

BD offers physicians valuable insights into urinary frequency, functional bladder capacity, most frequently observed bladder volume, urgency and urge urinary incontinence episodes, volume of water intake, nocturnal or daily urine output ratio, and associated bladder pain episodes [5]. They assess not only the storage phase but also some symptoms related to the voiding phase. For optimal adherence and reliability, a duration of 3-7 days is recommended for maintaining a BD [3,6]. Additionally,

BD is included in the guidelines of both the American Urological Association and the European Association of Urology for evaluating overactive bladder and LUTS [2-4].

Even though both of them can provide physicians with several parameters for evaluating the lower urinary tract, there are still some limitations of uroflowmetry and BD for depicting the whole picture of patients' LUTS, such as inconveniences, inadequate measurement frequency for uroflowmetry, lack of objective recording, and poor adherence for BD.

The aim of this paper is to discuss potential solutions for the earlier-mentioned limitations by reviewing (1) the evolution and current clinical use of uroflowmetry and BD, including clinical guidelines, daily practice applications, and their historical development; (2) a growing trend toward using home devices with various technologies; and (3) a comprehensive comparison of the strengths and weaknesses of these home devices.

Clinical Application and Limitation of Uroflowmetry and BD

Clinical Application and Limitation of Uroflowmetry

Uroflowmetry is typically performed in health care institutes, where patients may experience heightened emotional effects compared to when they are at home. Pre- and intratest anxiety (just like "white-coat hypertension") should be taken into consideration for its effect on lowering test reproducibility [7]. This practice also causes inadequate measurement frequency and a lack of voiding in different scenarios.

As to representativeness, some parameters are prone to within-subject variation [8-11], so it is recommended to repeat uroflowmetry measurements. The diagnostic accuracy of uroflowmetry is largely affected by threshold values [12,13], especially with physiological compensatory processes, detrusor underactivity, or an underfilled bladder [14]. Although uroflowmetry can be used to monitor treatment outcomes [15] and to correlate symptoms with these objective findings [12,16], its clinical value is still limited, as it is unable to differentiate between the possible underlying mechanisms. Again, specificity can be improved by repeated flow rate testing [6]. However, it is not feasible for a patient to receive multiple uroflowmetry in a clinical setting. Therefore, Caffarel et al [17] and Bray et al [18] even suggested the use of home uroflowmetry to achieve multiple measurements to improve the reliability of uroflowmetry.

In summary, uroflowmetry should be performed in a comfortable environment repeatedly to represent the whole picture of a patient's voiding pattern. This means that an at-home uroflowmetry is preferred.

Clinical Application and Limitation of BD

BD provides a method of quantifying symptoms, such as frequency of urge incontinence events and number of nocturia episodes [3], and reduces recall bias. However, it is not completely objective because the process relies on manual actions, such as recording the time and using measurement tools. These will affect BD's reliability.

Furthermore, the use of BD might induce a "bladder training" effect and influence the frequency of nocturnal voids [19]. Though it is important to guide patients to live an appropriate lifestyle [20], such modifications could potentially obscure the true micturition behavior of the patient. It is just like a clinical "uncertainty principle" (when you measure a system and change it at the same time), which should be taken into consideration when the goal of the test is to accurately depict the true picture of a patient's micturition. It is preferred to record a BD intuitively and insensibly.

The duration of BD has to be long enough to reduce sampling errors but short enough to avoid nonadherence [21]. It is recommended to conduct a BD including at least 3 days (continuous or separated) [22]. However, a longer BD is generally more reliable [23-26]. Because of the considerable effort required to complete each entry in a BD, there has been minimal achievement in comparing and validating BDs. Until now, the International Consultation on Incontinence Questionnaire BD is the only one that has undergone full validation [27]. The current protocol of BD necessitates that patients physically collect and measure each void, recording the volume along with associated symptoms. This practice requires patients' thorough intellectual understanding of the whole procedure and objectively recording BD. Among a survey of urogynecology or female pelvic medicine and reconstructive surgery specialists' fellowship-trained attendings, 25% had never or rarely (frequency of use <20%) used BD, and 97.5% reported difficulties associated with obtaining correctly completed and clinically applicable BD. The authors mentioned the best way is to teach patients with thorough instruction [20]. Therefore, a BD system capable of automation, objective recording, and repeated measurement appears to be the ideal solution.

Innovative and Possible Solutions for Avoiding Limitations

Evolution History of Uroflowmetry

As discussed in the "Clinical Application and Limitation of Uroflowmetry" section, it is important to provide patients with a comfortable environment and repeated measurements. In a clinical setting, urine weight–based, dipstick, and spinning-disk uroflowmetry are currently the most common conventional measurements of flow rate. Urine weight–based uroflowmetry detects the rate of changed weight or volume of voided urine. Dipstick uroflowmetry uses a dipstick immersed in voided urine sensing the change of fluid height and obtaining the flow rate. Spinning-disk uroflowmetry relates the power needed to counteract the slowing speed of the spinning disk hit by falling urine to the rate of urine flow [28]. In this paper, some trials of home uroflowmetry (relatively comfortable environment and feasible for repeated measurements) will be presented and discussed.

There have been several types of funnel-based uroflowmetry in the market or research fields [29-36]. The principle of funnel-based uroflowmetry is a combination of a funnel and some calibrating assembly for measuring Qmax. They all need

to be operated manually, and the results are then recorded on paper. One funnel-based home uroflowmetry has been developed for decades [37]. P-Flow (Tejnaksh Healthcare Ltd) is now a commercially available home uroflowmetry, which provides not only Qmax, average flow, voiding time of urine, flow curve, and total voided volume of urine but also simple urinalysis by an accompanied dipstick. However, users need to upload photos of the indicators on the device after finishing the test onto the health care company's server for interpreting the final results. There would be a time lag before obtaining the interpretation from the experts of this company. Additionally, it was found that Tejnaksh, the company behind this product, is a recognized teaching institute by the Ministry of Health and Family Welfare in India. To our knowledge, however, the product itself has not obtained certification from any other health authority.

There are at least 2 series of active research groups devoted to sound-based uroflowmetry, one from Singapore [38] and another from Korea [39]. Both systems use mobile phones to detect voiding sound and transfer it to the signal of voiding. Important urodynamic parameters, such as Qmax, voided volume, voiding time and average flow rate, and voiding flow patterns, can be presented through the embedded artificial intelligence (AI) algorithm. ProudP is currently commercially operated in the United States, which not only enhances the communication between patients and doctors but also improves the care quality for patients with LUTS. However, sound sensing uroflowmetry may be interfered in some noisy environments and even by the material of the testing urinal. There will be more discussion about the features of home devices in the "Comparison Between Home Devices of Different Technologies" section.

To provide a comfortable and intuitive environment for uroflowmetry, a toilet-shaped uroflowmetry (UM-100, Toto Ltd) was commercially developed in 2008 and was initially designed for clinical use [40]. A water-level measurement unit is installed with a connection to the toilet bowl. By balancing the hydraulic pressure between the measurement unit and the toilet bowl (communicating tubes in physics), urination-caused change in the water level can be measured and transformed to uroflowmetry parameters. The merit of this system is no necessity to do any cleaning work after each measurement. However, since the system detects the change in the total volume, it may have erroneous measurements when patients pass stool and urinate simultaneously [41].

A novel technology of vibration-based uroflowmetry was developed by a Taiwan-based interdisciplinary team [42]. This system simultaneously detects vibration signals during urination using an accelerometer alongside conventional uroflowmetry. Strong correlations were observed between this system and conventional uroflowmetry for parameters such as Qmax, voided volume, voiding time, and time to Qmax. Additionally, an AI model was used to analyze and predict 6 predefined patterns of uroflow curves, aiding in diagnosing voiding dysfunction with an accuracy of approximately 98%. This relatively low-cost system is suitable for automatic home urinary monitoring and enables repeated uroflow monitoring of patients outside health care institutions.

Evolution History of BD

With the reflection in the "Clinical Application and Limitation of BD" section, it is suggested to provide patients with automation, objective recording, and repeated measurements. It can be traced back to 1993 when the first computerized voiding diary "Compu-Void" was developed [43]. Compu-Void was a 64,000 RAM–capacity manual unit operated on the primitive operating system "DOS" of a personal computer. In the following decades, numerous trials have been conducted in pursuit of enhancing BD, aiming for automation, objective recording, and repeated measurement.

Quinn et al [44] ever developed a logical flow for asking users' symptoms on the portable electronic BD "MiniDoc." The advantage of an electronic data-inputting device lies in its accuracy and speed in retrieving data for further analysis [44]. Mangera et al [45] compared a paper-card reader and a hand-held input device with conventional written BD and found that an intuitive and user-friendly interface led to not only patients' preference but also the efficiency and accuracy of data management. With the onset of the digital era, application software (apps) on mobile phones and tablet computers began to play a crucial role in recording BD [46], and their numbers surged rapidly. Reports indicated that there were 55 apps available in languages such as Portuguese, Spanish, French, or English, with some offering functions for analyzing incontinence episodes and nocturia [47]. However, all of these apps require manual input.

For automated entry of data, Takai et al [48] ever introduced a body weight–based BD, which demonstrated a strong correlation between the voided urine weight recorded by the device and voided urine weight measured manually by the examinee [48]. This is the best practice of frequency or volume record and should be combined with entering episodes of LUTS to accomplish a comprehensive BD, though it cannot measure accurate urine weight during defecation.

With robust procedures of BD, the most common way to teach patients how to complete a BD is by providing detailed instructions along with any type of BD. In the current digital age, it should be expected that more clinicians like to use mobile apps or other digital resources. However, it was shown that very few clinicians actually used mobile apps (0.9%) or directed patients to use internet resources (1.2%) [20]. Furthermore, there is still no evidence to suggest that these digital apps can independently improve return rates or accuracy of the BD. In our opinion, the reason behind this is that there is only little difference between paper-based and electronic-based BDs. Electronic BDs simply record data through manual input or touch-screen processes, lacking automation. Moreover, the most cumbersome step of BD, collecting and measuring urine volume, remains unchanged. In this regard, a body weight–based BD offers certain advantages [48].

Since some important characteristics of a reliable BD, such as automation, objective recording, and repeated measurement, are fulfilled by the earlier-mentioned technologies (Table 1), we can envision a more advanced tool for accurately depicting a patient's voiding patterns. Additional suggestions for the development of a combined home uroflowmetry and BD system

are as follows: elimination of the need for cleaning after each use, incorporation of an intuitive and user-friendly interface, integration of wireless communication capability, accessibility of data for both users and doctors, and suitability for both storage and voiding phases. Therefore, such a comprehensive and intuitive "voiding recorder" will be a new standard for at-home voiding monitoring.

Table 1. Chronology of uroflowmetry and bladder diary (BD) advanced technologies in the recent century.

Year	Event	Remarks
1932	Ballenger suggested the maximum distance of a man's urine ejecting	[49]
1948	Drake recorded change in urine weight with time and manually calculated flow rate	[49,50]
1957	Kaufman improved Drake's system with electrical apparatus	[51]
1957	Von Garrelts electrically calculated flow rate	[41]
1965	Smith designed funnel-based uroflowmetry	[29]
1976	Drach used a dipstick to estimate the flow rate	[35]
1993	The first computerized voiding diary "Compu-Void" was developed	[43]
1999	International Continence Society defined uroflowmetry parameters	[52]
2003	Quinn drew a logical flow for asking users' symptoms	[44]
2008	Toto developed communicating tube uroflowmetry	[40]
2014	Mangera used paper-card reader	[45]
2014	Bright designed the first standardized and validated BD (ICIQ[a] BD)	[27]
2015	Krhut compared uroflowmetry and sonouroflowmetry	[53]
2016	Application of mobile phones and tablet computers in recording BD	[46]
2021	Takai introduced a body weight–based BD	[48]
2022	Pong linked vibration with uroflowmetry	[42]

[a]ICIQ: International Consultation on Incontinence Questionnaire.

The Emerging Trend of Using Home Devices

As we enter the era of AI, big data become crucial for training and applying AI models. This enhances the value of home devices, particularly for their ability to repeatedly measure home uroflowmetry and generate big data. Home devices facilitate mobile health or medicine by offering a comfortable environment, enabling repeated measurements, and providing big data for communication and application. It has been noted that extensive at-home data are often more reliable than single in-office tests [54]. Recently, there has been a surge in interest in home uroflowmetry at international scientific meetings and in literature [9,55-58]. Several commercial home uroflowmetry devices have also been introduced, offering repeated measurements of voiding parameters and uroflow curves. This method of repeated home measurement can provide a comprehensive picture of patients' daily-life voiding without the anxiety and stress associated with office uroflowmetry tests [9,55]. As mentioned in the "Introduction" section, uroflowmetry primarily describes the condition of the voiding phase, while BD addresses the storage phase and some symptoms related to voiding. Each can be applied to patients with different types of voiding dysfunction. Following advancements, home devices can now measure parameters in both the storage and voiding phases, making them suitable for almost all patients with voiding dysfunction.

Certainly, there is a need for home uroflowmetry to offer repeated measurements and convenience, addressing several clinical and practical challenges for practicing urologists [54]. With various home uroflowmetry technologies emerging in the market, it is important to understand their strengths and weaknesses. What are the comparative advantages and disadvantages of these different home uroflowmetry technologies?

Comparison Between Home Devices of Different Technologies

All types of home uroflowmetry technologies share common strengths, such as automatic recording, repeated measurements, and the generation of big data. However, each technology also has its own unique strengths and weaknesses when used for home uroflowmetry. Table 2 outlines various features of these technologies, including accuracy, susceptibility to environmental interferences, uroflow curve pattern recognition, AI algorithms, and contact-free operation.

Table 2. The comparison on different technologies for home uroflowmetry.

Features or technologies	Weighing (gravimetric)	Height sensor	Sound	Vibration
Accuracy	FDA[a] certified [57]	FDA certified [59]	Correlation with office uroflowmetry (R=0.91) [54]; prediction rate of 99% [60]	Uroflow curve pattern recognition accuracy>0.98 [42]
Vulnerability to the surrounding interferences	No	No	Yes [61]	No
Uroflow curve pattern recognition	No	No	Yes [60]	Yes
AI[b] algorithm or model	No	No	Yes [60,61]	Yes
Contact-free (no need for installation or cleaning)	No	No	Yes	Yes

[a]FDA: Food and Drug Administration.

[b]AI: artificial intelligence.

In terms of accuracy, all technologies yield results comparable to office uroflowmetry, with some even receiving Food and Drug Administration approval. However, sound-based technologies are sensitive to environmental interferences like noise or barriers due to their measuring mechanism. For accurate recording, the urine stream must be voided into a water-filled commode rather than a urinal [54]. Additionally, the sound of voiding varies between men and women, likely due to anatomical and postural differences during urination. Men, who typically urinate standing up, produce louder sounds, as there are no barriers to dampen the sound. In contrast, women, who usually urinate sitting down, produce quieter sounds, and the sitting position can block sound transmission. To address these issues, different models have been developed for both genders [61].

Vibration-based technology offers a potential solution by reducing the barriers to sound transmission. Vibration is transmitted more directly than sound, which travels through the air, while vibration is conveyed through the concrete toilet bowl or urinal to the sensor. Consequently, sound-based technologies are more vulnerable to surrounding noise compared to vibration-based ones [56].

After reviewing the literature, it is evident that only sound- and vibration-based technologies have used AI models or algorithms for uroflow curve pattern recognition [42,60,61]. These 2 technologies are advantageous for intuitive and contact-free measurements, as they do not require installation or cleaning for each use. The calibration of any device is crucial and must be verified during the approval process. It should be ensured that all home devices receive approval from health authorities before clinical application. Additionally, with repeated measurements using home devices, the current reference ranges for clinical urodynamic studies may be revised accordingly. However, such revisions should be undertaken only after sufficient evidence has been accumulated. These home devices not only enhance accessibility for patients and physicians in assessing voiding patterns, but they also improve the diagnosis and treatment process by collecting more comprehensive information. The use of home devices is anticipated to transform the management of LUTS.

Conclusions

Uroflowmetry and BD are key diagnostic tools for LUTS, despite facing limitations such as the absence of a comfortable environment and infrequent measurements for uroflowmetry as well as the lack of automation, objective recording, and repeated measurement for BD. Technological advancements have addressed some of these limitations.

In our opinion, the following points can be highlighted: (1) the emerging trend of using home devices can enhance diagnostic capabilities through repeated measurements and the convenience of at-home testing and (2) home devices, which provide both frequency-volume and uroflowmetry information, have the potential to transform the management of LUTS.

Acknowledgments

Microsoft Copilot was used for assisting wording and grammar polishing in the manuscript. The study received funding from Taipei Tzu-Chi Hospital, Buddhist Tzu-Chi Medical Foundation (TCRD-TPE-113-66).

Authors' Contributions

VFST, YHP, and YTT were responsible for the conceptualization of the study. YCT, SSDY, and YTT provided the necessary resources and supervised the project. VFST and YCT conducted the validation. VFST and MWL wrote the original draft. YCT, SSDY, YHP, YTT, and VFST reviewed and edited the manuscript.

Conflicts of Interest

None declared.

References

1. Rosier PFWM, Schaefer W, Lose G, Goldman HB, Guralnick M, Eustice S, et al. International Continence Society Good Urodynamic Practices and Terms 2016: urodynamics, uroflowmetry, cystometry, and pressure-flow study. Neurourol Urodyn 2017;36(5):1243-1260. [doi: 10.1002/nau.23124] [Medline: 27917521]
2. Sandhu JS, Bixler BR, Dahm P, Goueli R, Kirkby E, Stoffel JT, et al. Management of lower urinary tract symptoms attributed to benign prostatic hyperplasia (BPH): AUA Guideline amendment 2023. J Urol 2024;211(1):11-19. [doi: 10.1097/ju.0000000000003698]
3. Non-neurogenic female LUTS. European Association of Urology. 2024. URL: https://uroweb.org/guidelines/non-neurogenic-female-luts/chapter/citation-information [accessed 2024-12-28]
4. Management of non-neurogenic male LUTS. European Association of Urology. 2024. URL: https://uroweb.org/guidelines/management-of-non-neurogenic-male-luts/chapter/citation-information [accessed 2024-12-28]
5. Kuo HC. Interpreting the voiding diary of patients with lower urinary tract symptoms. Incont Pelvic Floor Dysfunct 2010;4(4):105-110.
6. ACOG. Practice Bulletin No. 155: urinary incontinence in women. Obstet Gynecol 2015;126(5):e66-e81. [doi: 10.1097/aog.0000000000001148]
7. Rubilotta E, Balzarro M, Castellani D, Tiso L, Panunzio A, Pirola GM, et al. The role of emotional condition in patients with lower urinary tract symptoms performing uroflowmetry. Urology 2021;148:37-46. [doi: 10.1016/j.urology.2020.10.045] [Medline: 33217454]
8. Golomb J, Lindner A, Siegel Y, Korczak D. Variability and circadian changes in home uroflowmetry in patients with benign prostatic hyperplasia compared to normal controls. J Urol 1992;147(4):1044-1047. [doi: 10.1016/s0022-5347(17)37462-1] [Medline: 1372659]
9. Summers SJ, Armstrong JM, Kaplan SA, Te AE, Le A, Heiner SM, et al. Male voiding behavior: insight from 19,824 at-home uroflow profiles. J Urol 2021;205(4):1126-1132. [doi: 10.1097/ju.0000000000001504]
10. Jørgensen JB, Jensen KM, Mogensen P. Age-related variation in urinary flow variables and flow curve patterns in elderly males. Br J Urol 1992;69(3):265-271. [doi: 10.1111/j.1464-410x.1992.tb15526.x] [Medline: 1373664]
11. Kranse R, van Mastrigt R. Causes for variability in repeated pressure-flow measurements. Urology 2003;61(5):930-934. [doi: 10.1016/s0090-4295(03)00010-4] [Medline: 12736007]
12. Reynard JM, Yang O, Donovan JL, Peters TL, Schafer W, de la Rosette JJ, et al. The ICS-'BPH' Study: uroflowmetry, lower urinary tract symptoms and bladder outlet obstruction. Br J Urol 1998;82(5):619-623. [doi: 10.1046/j.1464-410x.1998.00813.x] [Medline: 9839573]
13. Idzenga T, Pel JJ, van Mastrigt R. Accuracy of maximum flow rate for diagnosing bladder outlet obstruction can be estimated from the ICS nomogram. Neurourol Urodyn 2008;27(1):97-98. [doi: 10.1002/nau.20480] [Medline: 17600368]
14. Siroky MB, Olsson CA, Krane RJ. The flow rate nomogram: I. Development. J Urol 1979;122(5):665-668. [doi: 10.1016/s0022-5347(17)56550-7] [Medline: 159366]
15. Siroky MB, Olsson CA, Krane RJ. The flow rate nomogram: II. Clinical correlation. J Urol 1980;123(2):208-210. [doi: 10.1016/s0022-5347(17)55859-0] [Medline: 7354519]
16. Reynard JM, Peters TJ, Lim C, Abrams P. The value of multiple free-flow studies in men with lower urinary tract symptoms. Br J Urol 1996;77(6):813-818. [doi: 10.1046/j.1464-410x.1996.00097.x] [Medline: 8705213]
17. Caffarel J, Robson W, Pickard R, Griffiths C, Drinnan M. Flow measurements: can several "wrongs" make a "right"? Neurourol Urodyn 2007;26(4):474-480. [doi: 10.1002/nau.20383] [Medline: 17245778]
18. Bray A, Griffiths C, Drinnan M, Pickard R. Methods and value of home uroflowmetry in the assessment of men with lower urinary tract symptoms: a literature review. Neurourol Urodyn 2012;31(1):7-12. [doi: 10.1002/nau.21197] [Medline: 22038511]
19. Vaughan CP, Johnson TM, Goode PS, Redden DT, Burgio KL, Markland AD. Military exposure and urinary incontinence among American men. J Urol 2014;191(1):125-129. [doi: 10.1016/j.juro.2013.07.016] [Medline: 23871759]
20. Mehta S, Geng B, Xu X, Harmanli O. Current state of bladder diary: a survey and review of the literature. Int Urogynecol J 2023;34(4):809-823. [doi: 10.1007/s00192-022-05398-w] [Medline: 36322174]
21. Bright E, Drake MJ, Abrams P. Urinary diaries: evidence for the development and validation of diary content, format, and duration. Neurourol Urodyn 2011;30(3):348-352. [doi: 10.1002/nau.20994] [Medline: 21284023]
22. Yap TL, Cromwell DC, Emberton M. A systematic review of the reliability of frequency‐volume charts in urological research and its implications for the optimum chart duration. BJU Int 2006;99(1):9-16. [doi: 10.1111/j.1464-410x.2006.06499.x]
23. Dmochowski RR, Sanders SW, Appell RA, Nitti VW, Davila GW. Bladder‐health diaries: an assessment of 3‐day vs 7‐day entries. BJU Int 2005;96(7):1049-1054. [doi: 10.1111/j.1464-410x.2005.05785.x]
24. Schick E, Jolivet-Tremblay M, Dupont C, Bertrand PE, Tessier J. Frequency-volume chart: the minimum number of days required to obtain reliable results. Neurourol Urodyn 2003;22(2):92-96. [doi: 10.1002/nau.10079] [Medline: 12579624]
25. Groutz A, Blaivas JG, Chaikin DC, Resnick NM, Engleman K, Anzalone D, et al. Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: a multicenter study of micturition diary and pad tests. J Urol 2000;164(3 Part 1):698-701. [doi: 10.1016/s0022-5347(05)67284-9]

26. Dixon CA, Nakib NA. Are bladder diaries helpful in management of overactive bladder? Curr Bladder Dysfunct Rep 2016;11(1):14-17. [doi: 10.1007/s11884-016-0343-x]
27. Bright E, Cotterill N, Drake M, Abrams P. Developing and validating the International Consultation on Incontinence Questionnaire bladder diary. Eur Urol 2014;66(2):294-300. [doi: 10.1016/j.eururo.2014.02.057] [Medline: 24647230]
28. Addla SK, Marri RR, Daayana SL, Irwin P. Avoid cruising on the uroflowmeter: evaluation of cruising artifact on spinning disc flowmeters in an experimental setup. Neurourol Urodyn 2010;29(7):1301-1305. [doi: 10.1002/nau.20846] [Medline: 20879000]
29. Smith JC. An individual uroflowmeter. Lancet 1965;1(7376):90. [doi: 10.1016/s0140-6736(65)91667-3] [Medline: 14234215]
30. Currie RJ. The Streamtest cup: a new uroflow device. Urology 1998;52(6):1118-1121. [doi: 10.1016/s0090-4295(98)00404-x] [Medline: 9836566]
31. Pel JJM, van Mastrigt R. Development of a low-cost flow meter to grade the maximum flow rate. Neurourol Urodyn 2002;21(1):48-54. [doi: 10.1002/nau.2112] [Medline: 11835424]
32. Pridgeon S, Harding C, Newton D, Pickard R. Clinical evaluation of a simple uroflowmeter for categorization of maximum urinary flow rate. Indian J Urol 2007;23(2):114-118 [FREE Full text] [doi: 10.4103/0970-1591.32058] [Medline: 19675784]
33. Ball AJ. The peakometer: an evaluation. Urol Int 1982;37(1):42-44. [doi: 10.1159/000280794] [Medline: 6214070]
34. Colstrup H, Andersen JT, Walter S, Rytt S. A disposable uroflowmeter for recording of maximum flow rate. Accuracy and clinical acceptability. Scand J Urol Nephrol 2010;17(3):303-306. [doi: 10.3109/00365598309182136]
35. Drach GW, Binard W. Disposable peak urinary flowmeter estimates lower urinary tract obstruction. J Urol 1976;115(2):175-179. [doi: 10.1016/s0022-5347(17)59122-3] [Medline: 129576]
36. Lucas M, Moore K, Borkowski T, Pantuck A, Heesakkers J, Piotr P. Accuracy of CaptiFlow portable home flow measuring device and its use as a diagnostic tool in the management of LUTS—a randomised cross over study. 2009 Presented at: ICS Annual Meeting; 2009; San Francisco, CA, United States URL: https://www.ics.org/2009/abstract/400
37. De La Rosette JJ, Witjes WP, Debruyne FM, Kersten PL, Wijkstra H. Improved reliability of uroflowmetry investigations: results of a portable home-based uroflowmetry study. Br J Urol 1996;78(3):385-390 [FREE Full text] [doi: 10.1046/j.1464-410x.1996.00115.x] [Medline: 8881947]
38. Lee HJ, Aslim EJ, Balamurali B, Ng LYS, Kuo TLC, Lin CMY, et al. Development and validation of a deep learning system for sound-based prediction of urinary flow. Eur Urol Focus 2023;9(1):209-215. [doi: 10.1016/j.euf.2022.06.011] [Medline: 35835694]
39. Kim H, Ye C, Jung G, Ryu H, Lee JW, Jeong SJ, et al. Validation of acoustic voided volume measure: a pilot prospective study. World J Urol 2023;41(2):509-514. [doi: 10.1007/s00345-022-04231-9] [Medline: 36550234]
40. Izumi H, Nakagawa H, Kaiho Y, Miyazato M, Kawamorita N, Arai Y. Difference of uroflow parameters between novel toilet-shaped uroflowmeter (water level sensor mechanism) and conventional uroflowmeter in urologically normal female patients. 2011 Presented at: International Continence Society Annual Meeting; August 29-September 2, 2011; Glasgow.
41. Chun KS, Kim SJ, Cho ST. Noninvasive medical tools for evaluating voiding pattern in real life. Int Neurourol J 2017;21(Suppl 1):S10-S16. [doi: 10.5213/inj.1734860.430] [Medline: 28446014]
42. Pong YH, Tsai VFS, Hsu YH, Lee CH, Wang KC, Tsai YT. Application of a deep learning neural network for voiding dysfunction diagnosis using a vibration sensor. Appl Sci 2022;12(14):7216. [doi: 10.3390/app12147216]
43. Rabin JM, McNett J, Badlani GH. Computerized voiding diary. Neurourol Urodyn 1993;12(6):541-553. [doi: 10.1002/nau.1930120604] [Medline: 8312939]
44. Quinn P, Goka J, Richardson H. Assessment of an electronic daily diary in patients with overactive bladder. BJU Int 2003;91(7):647-652. [doi: 10.1046/j.1464-410x.2003.04168.x] [Medline: 12699477]
45. Mangera A, Marzo A, Heron N, Fernando D, Hameed K, Soliman AA, et al. Development of two electronic bladder diaries: a patient and healthcare professionals pilot study. Neurourol Urodyn 2014;33(7):1101-1109. [doi: 10.1002/nau.22469] [Medline: 24000163]
46. Sussman RD, Richter LA, Tefera E, Park AJ, Sokol AI, Gutman RE, et al. Utilizing technology in assessment of lower urinary tract symptoms: a randomized trial of electronic versus paper voiding diaries. Female Pelvic Med Reconstr Surg 2016;22(4):224-228. [doi: 10.1097/SPV.0000000000000266] [Medline: 27054797]
47. Vaccari NA, da Silveira LTY, Bortolini MAT, Haddad JM, Baracat EC, Ferreira EAG. Content and functionality features of voiding diary applications for mobile devices in Brazil: a descriptive analysis. Int Urogynecol J 2020;31(12):2573-2581. [doi: 10.1007/s00192-020-04382-6] [Medline: 32592019]
48. Takai S, Matsukawa Y, Hashizume N, Gotoh M. A small pilot study to evaluate the accuracy and feasibility of a novel automated voiding diary device for recording urine output measurements. Neurourol Urodyn 2021;40(1):272-277. [doi: 10.1002/nau.24547] [Medline: 33137844]
49. Drake WM. The uroflometer: an aid to the study of the lower urinary tract. J Urol 1948;59(4):650-658. [doi: 10.1016/s0022-5347(17)69425-4]
50. Chancellor MB, Rivas DA, Mulholland S, Drake WM. The invention of the modern uroflowmeter by Willard M. Drake, Jr at Jefferson Medical College. Urology 1998;51(4):671-674. [doi: 10.1016/s0090-4295(97)00203-3] [Medline: 9586627]
51. Kaufman JJ. A new recording uroflometer: a simple automatic device for measuring voiding velocity. J Urol 1957;78(1):97-102. [doi: 10.1016/s0022-5347(17)66405-x]

52. Stöhrer M, Goepel M, Kondo A, Kramer G, Madersbacher H, Millard R, et al. The standardization of terminology in neurogenic lower urinary tract dysfunction: with suggestions for diagnostic procedures. International Continence Society Standardization Committee. Neurourol Urodyn 1999;18(2):139-158. [doi: 10.1002/(sici)1520-6777(1999)18:2<139::aid-nau9>3.0.co;2-u] [Medline: 10081953]
53. Krhut J, Gärtner M, Sýkora R, Hurtík P, Burda M, Luňáček L, et al. Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men. Int J Urol 2015;22(8):761-765. [doi: 10.1111/iju.12796] [Medline: 25988672]
54. Schultz RE. Smartphone app for in-home uroflowmetry. Urol Pract 2022;9(6):524-530. [doi: 10.1097/upj.0000000000000338]
55. Bladt L, Kashtiara A, Platteau W, De Wachter S, De Win G. First-year experience of managing urology patients with home uroflowmetry: descriptive retrospective analysis. JMIR Form Res 2023;7:e51019 [FREE Full text] [doi: 10.2196/51019] [Medline: 37847531]
56. Kumar S, Kumar M. MP09-09 mobile app-based versus conventional uroflowmetry: is your home toilet the new uroflowmetry lab. J Urol 2024;211(5S):211-229. [doi: 10.1097/01.ju.0001008920.55771.18.09]
57. Morales EVB, Peters M, Pauwels J, Vermandel A, Wachter SD, Bladt L, et al. MP75-13 digital health solutions create opportunities for home uroflowmetry. J Urol 2024;211(5S):e1237. [doi: 10.1097/01.ju.0001008676.21744.5f.13]
58. Song SH, Park JH, Park HJ, Jeong Y, Ryu H, Lee JW, et al. MP09-12 a prospective multicenter clinical trial for efficacy of mobile uroflowmetry in treatment-naive benign prostatic hyperplasia undergoing medical therapy: interim results. J Urol 2024;211(5S):e131. [doi: 10.1097/01.ju.0001008920.55771.18.12]
59. Patel D, Ogden K, Anderl W, Mastrangelo S, Le A, Hotaling J, et al. PD50-07 non-inferiority of a novel capacitance fluid height design compared to rotating disc and weight transducer designs for home uroflowmetry applications. J Urol 2018;199(4S):e972. [doi: 10.1016/j.juro.2018.02.2335]
60. Jin J, Chung Y, Kim W, Heo Y, Jeon J, Hoh J, et al. Classification of bladder emptying patterns by LSTM neural network trained using acoustic signatures. Sensors (Basel) 2021;21(16):5328 [FREE Full text] [doi: 10.3390/s21165328] [Medline: 34450769]
61. Lee YJ, Kim MM, Song SH, Lee S. A novel mobile acoustic uroflowmetry: comparison with contemporary uroflowmetry. Int Neurourol J 2021;25(2):150-156 [FREE Full text] [doi: 10.5213/inj.2040250.125] [Medline: 33387990]

Abbreviations

AI: artificial intelligence
BD: bladder diary
LUTS: lower urinary tract symptom
Qmax: maximum flow rate

Edited by T de Azevedo Cardoso; submitted 20.09.24; peer-reviewed by A Gammie, A Rantell; comments to author 19.11.24; revised version received 24.11.24; accepted 29.11.24; published 28.01.25.

Please cite as:
Li MW, Tsai YC, Yang SSD, Pong YH, Tsai YT, Tsai VFS
The Evolution of Uroflowmetry and Bladder Diary and the Emerging Trend of Using Home Devices From Hospital to Home
Interact J Med Res 2025;14:e66694
URL: https://www.i-jmr.org/2025/1/e66694
doi:10.2196/66694
PMID:

Viewpoint

The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations

Dimitri Roustan[1], MBBS; François Bastardot[2], MD, MSACI

[1]Emergency Medicine Department, Cliniques Universitaires Saint-Luc, Brussels, Belgium

[2]Medical Directorate, Lausanne University Hospital, Lausanne, Switzerland

Corresponding Author:
Dimitri Roustan, MBBS
Emergency Medicine Department
Cliniques Universitaires Saint-Luc
Avenue Hippocrate 10
Brussels, 1200
Belgium
Phone: 32 477063174
Email: dim.roustan@gmail.com

Abstract

Large language models (LLMs) are artificial intelligence tools that have the prospect of profoundly changing how we practice all aspects of medicine. Considering the incredible potential of LLMs in medicine and the interest of many health care stakeholders for implementation into routine practice, it is therefore essential that clinicians be aware of the basic risks associated with the use of these models. Namely, a significant risk associated with the use of LLMs is their potential to create hallucinations. Hallucinations (false information) generated by LLMs arise from a multitude of causes, including both factors related to the training dataset as well as their auto-regressive nature. The implications for clinical practice range from the generation of inaccurate diagnostic and therapeutic information to the reinforcement of flawed diagnostic reasoning pathways, as well as a lack of reliability if not used properly. To reduce this risk, we developed a general technical framework for approaching LLMs in general clinical practice, as well as for implementation on a larger institutional scale.

(Interact J Med Res 2025;14:e59823) doi:10.2196/59823

KEYWORDS

medical informatics; large language model; clinical informatics; decision-making; computer assisted; decision support techniques; decision support; decision; AI; artificial intelligence; artificial intelligence tool; LLM; electronic data system; hallucinations; false information; technical framework

Introduction to Large Language Models

The development of artificial intelligence (AI) solutions and their recent democratization have allowed the public to access various innovative tools. Notably, several large language models (LLMs) have recently surged in popularity due to significant media attention and by offering free access for registered users (eg, ChatGPT, Gemini, and Meta LLaMA).

An LLM is a type of deep learning model that is pretrained on large text datasets. They are often based on the transformer architecture [1], an innovative form of neural network that uses an encoder-decoder structure to rapidly process large blocks of text, avoiding redundancies that hampered recurrent neural networks in the past. Several popular LLMs have integrated a chatbot interface to allow users to interact directly with the model, generating appropriate, context-aware responses to a user's input in a conversational manner. This allows the user to engage in dynamic conversations that appear natural, making the technology a powerful tool for various applications across a wide range of fields.

The use of LLMs has become widespread, and medicine is no exception. LLMs have the potential of becoming a disruptive tool in medicine [2] and will certainly have a major impact on clinical practice, medical education, and research. Within this field, LLM performance has already been evaluated to take specialist board exams [3], improve communication with patients [4,5], and write drafts for scientific papers. It can also be an interesting tool for creating a dynamic learning experience [6]. Although LLMs with chatbot interfaces represent an interesting tool for many tasks in clinical medicine and research, users should be aware of one of the most significant shortcomings of these models, called "hallucinations."

In this viewpoint, we review the underlying causes of hallucinations in LLMs and examine their implications within the field of clinical medicine. We also explore current and future strategies for mitigating these limitations and present a general framework to guide clinicians in critically assessing and integrating LLMs into clinical practice.

Overview of Hallucinations

In medicine, hallucinations refer to sensory experiences that occur in the absence of corresponding external stimuli. In the field of LLMs, hallucinations refer to the generation of false or fabricated information. This signifies that the LLM will create nonfactual content to answer a user's question without clarifying whether the answer contains fabricated information. Hallucinations stem from many root causes, which we will delve into below.

First, both the quality and volume of the dataset upon which the LLM has been trained are important variables and can explain the number of hallucinations the LLM produces to some degree [7,8]. How data are collected and how the model is trained can also influence hallucination frequency [8]. Furthermore, the method through which the editor fine-tunes the model can also influence the final output.

Another major cause of hallucinations stems from the very way certain LLMs are programmed. Indeed, most LLMs are auto-regressive; the term "auto-regressive" refers to the model's ability to predict future elements of a sequence based on its previous outputs. These elements, usually one or multiple words, are termed tokens. An auto-regressive LLM aims to produce an output based on token prediction; this signifies that the model will predict the most probable next token(s) given a specific input token. In practice, it predicts the following word(s) after the sequence of words it has already given. However, it generates each next token by considering the previous one, and not the whole sequence. This means words are generated in a word-after-word fashion, without necessarily using the whole of the previously generated sentence to predict the rest of the sentence [9]. This lays the groundwork for producing hallucinations since factual accuracy is not the end goal. Rather, accuracy is inferred from a high probability of adequate token prediction based on the data in the training dataset. Since the dataset is necessarily flawed or incomplete, hallucinations can arise.

The size of the training dataset can also influence hallucination type and degree. It has been demonstrated on multiple LLMs that the larger the training dataset size, the more likely the model will be capable of recognizing its limitations and acknowledging uncertainty [10]. Furthermore, choices made by the editor will also influence output quality (ie, fine-tuning decisions, output ranking, censorship, etc).

User input is also of great importance in determining the quality of the model output. Indeed, it has been shown that user input through contextualization and inclusion of source material can also modify the number of hallucinations an LLM produces [11].

Implications for Clinical Medicine

LLMs have many potential benefits in the health care system, for both providers and patients. For simple tasks, it is highly likely that part or all of the process will be carried out with LLM tools in the foreseeable future. Efficiency will likely be improved by reducing redundant and tedious tasks (most likely administrative before clinical) [12-14], and there may even be applications for reducing diagnostic delay for difficult diagnoses [15]. Nonetheless, despite these potential benefits, hallucinations represent a major risk if unaccounted for when using LLMs [16-18]. Below, we will review some situations that have appeared apparent to us when testing LLMs.

In practice, LLMs may erroneously attribute clinical, biological, or radiological characteristics to certain diseases or conditions, depending on the way the clinician inputs data as well as the probabilistic behavior of the model. This flaw, in combination with anchoring and confirmation bias, may unknowingly lead the clinician down an erroneous diagnostic or therapeutic pathway. This can have severe consequences for the patient's health.

LLMs may also make false claims about diagnostic accuracy for diagnostic procedures. This can lead the clinician to either overestimate or underestimate the diagnostic capacity of a procedure. The consequences could be either depriving a patient of a reliable diagnostic method or, on the contrary, relying on an inadequate diagnostic method to make a statement about the disease process. More specifically in the latter case, the absence of a disease process may be wrongly inferred based on an insensitive exam, and the presence of a disease process may be improperly inferred from a nonspecific exam.

Furthermore, the LLM may suggest inadequate workups and therapeutic procedures. It is important to remember that LLMs are trained on databases that may either not encompass the data necessary to provide adequate guidance (ie, absence of medical guidelines) or contain outdated medical recommendations. Further, given the crucial importance of input data supplied by the user, the omission of a simple characteristic may cause the LLM to produce an inadequate plan of care. Moreover, the LLM may not necessarily prompt the user for additional information regarding important characteristics that could influence the plan of care; most notably, social characteristics and cultural preferences may be inadequately accounted for. Specific diagnostic or therapeutic measures proposed by the LLM may be inadequate or inappropriate, based on important parameters such as pretest probability, as well as patient preferences and prognosis. In addition, given the diversity of health care system models in countries around the world, the inherent bias introduced by the LLM's dataset can lead to recommending inadequate plans of care for a different health care model than that which the dataset contains information on.

Consistency, and thus reliability, is another issue that can appear while using LLMs to make recommendations for plans of care. Indeed, even with a consistently identical user query, the information contained within the LLMs response may vary considerably when the user renews the query. This variability is an important consideration when the clinician is contemplating

the possibility of integrating LLMs into the patient care process. Indeed, it is well accepted in the health care quality community that reducing variability to a certain degree is an essential step in increasing quality [19]; this is even more so true when some parts of the information provided by the LLM are at risk of being inadequate.

In addition, the use of LLMs in clinical practice raises a significant number of ethical concerns, which we only slightly touch upon in this viewpoint. Questions regarding overreliance on LLMs and other AI tools, as well as the legal and ethical ramifications of decision-making based upon AI input, are crucial. In the era of evidence-based medicine, LLM source material and information traceability will be essential in order to reliably inform patient care decisions. The value of the clinician's experience in nuancing the LLM's outputs will also remain critical in delivering personalized patient care.

To build upon the legal and ethical concerns related to these LLMs, it is important to keep in mind that many LLMs are developed by private companies and are not open source. Data management, especially related to patient privacy rights, is an essential concern related to information input into the LLM. Implementation of LLM components in electronic medical record programs is being considered and poses the same risks. Notably, the question stands on the use of personal data to for-profit ends (targeted advertising or selling health care data to insurance companies). Local institutional governance and national regulations will be essential in managing and mitigating these potential risks.

Finally, LLM use also carries considerable potential to alter the patient-clinician relationship. Patients may increasingly discuss their symptoms and conditions with LLMs before seeing a medical professional, in a similar way that some patients use search engines before consulting a physician today. This poses the risk of fostering misguided self-diagnoses, with potentially harmful health consequences, especially in settings where access to health care can be financially challenging. Furthermore, patients might develop unrealistic expectations or demand unnecessary clinical resources, based on the information the LLM has provided. As a result, this could affect the dynamics of the patient-clinician relationship. On the other hand, LLMs possess the capacity to tailor medical information to the patient's level of comprehension, which may help enhance therapeutic education and adherence to medical advice.

Mitigating Hallucinations in Clinical Practice

Hallucinations can therefore represent a significant source of error if unaccounted for when using LLMs. A proactive and systematic approach is necessary to help interpret LLM output data and avoid succumbing to avoidable pitfalls that could cause harm to patients. This approach is summarized in Textbox 1.

Textbox 1. Important technical considerations before integrating large language models (LLMs) into clinical practice.

- What dataset was the LLM trained upon?
  - What specific considerations does this entail, with regard to bias?
  - Is the dataset up to date?
- What organization is behind the LLM and the dataset?
- Is the LLM specifically tailored for medical purposes?
- In testing rounds:
  - Is the information given by the LLM consistent with the existing knowledge on the subject?
  - Are the recommendations made by the LLM adequate compared to the accepted standard of care?
- How much variability exists within the LLM's responses? Can it be clinically significant for patient care?
- Enhanced capacity:
  - Does the model possess the capacity to integrate up-to-date information?
  - Does the model possess the capacity to search within reliable sources of information to better respond to the user's request? If the answer to one of the two above questions is yes, is this feature integrated within the LLM or is it operated by a third-party plug-in?
    - If a third-party plug-in is involved, what strengths and shortcomings does it entail?
  - Does the LLM possess the capacity to assess its answers' reliability?
  - Is the LLM capable of providing the links to its sources of information?
- Do best practice guidelines exist for the utilization of the LLM?
  - If so, are they specific to use in the health care sector?
  - If the guidelines are general use or specifically focused on another industry sector, what precautions must be applied before extrapolating their use to the health care sector?
- Has the model been tested in a rigorous fashion, and are the results of this evaluation subject to scientific publication?

First and foremost, understanding the model's origins, version, training database content, and strengths as well as drawbacks are essential prerequisites for an informed use of the LLM. With this information, the user should actively seek out what types of bias the model may contain and understand how it can affect the LLM's answers [2]. Furthermore, information on the training dataset should be sought out to understand how up-to-date the knowledge within it is, as well as if it is well equipped to answer medical inquiries. In this regard, a topic of emerging importance is the development of LLMs specifically trained for medical purposes. Although theoretically more performant than general LLMs, their relevance for clinical practice has not yet been evaluated.

Second, user input should be carefully crafted to create a high-quality request. The request should contain a detailed description of the clinical context; this requires carrying out a thorough history, clinical exam, and incorporating current as well as historical workup data. Clinical acumen thus remains essential in creating an adequate request. Therefore, although initially time-consuming, a higher-quality request can yield a more relevant answer.

Third, model accuracy and hallucination prevalence should be assessed before being put into practice, through iterative testing and evaluation. During testing rounds, LLM accuracy should be examined using standardized scenarios. Consistency, as well as variability in the answers, should be evaluated by regenerating the LLM's responses multiple times. Whether through formal, statistical evaluation or through getting a general sense of the model's characteristics, the clinician can evaluate the LLM's capabilities and shortcomings in this manner. Furthermore, the scientific adequacy of the LLM's responses should be assessed with regard to current standards of care and up-to-date guidelines. Even without knowing the database's knowledge cutoff, this method can help assess how up-to-date the data are, as well as understand how often hallucinations arise with regard to a specific subject.

Another useful tool that can help the clinician assess the reliability of the LLM's responses is plug-ins. These are usually third-party apps that can be programmed to serve a wide range of functions, including but not limited to, searching the internet, retrieving information from scientific databases, and substantiating responses with links to the sources of the presented information. LLMs may also possess certain of these capacities directly within the scope of their own functions. Although plug-ins may significantly enhance the reliability of the LLM's responses, by providing the ability for up-to-date referencing, they are not a guarantee that the response will be free from hallucinations. Therefore, plug-ins should be evaluated with the same amount of scrutiny as the LLM itself.

Finally, combining text, image, and video data in LLM training databases can lead to more accurate responses and may decrease the likelihood of hallucinations [20]. However, it is also important to remember that the multimodal model's performances still rely on the quality of their training dataset [21].

A crucial aspect in the deployment and judicious implementation of LLMs in a clinical setting lies in the establishment of a proactive error reporting program. In conjunction with the aforementioned recommendations, the implementation of such a program facilitates the identification and reporting of near-miss incidents. At the individual level, this allows the user to develop a personal appreciation of the model's shortcomings, as well as the topics subject to hallucinations. On an institutional level, it can help develop best practice guidelines by identifying frequent hallucination presentations and more general errors. If LLM solutions are delivered as an on-premises solution, it is conceivable that error reporting will help fine-tune local models.

Prospects

Research in AI will largely contribute to reducing hallucinations, be it through fine-tuning of the underlying model, prompt engineering techniques, development of specific medical LLMs, or other innovative approaches.

Given the increasing awareness of hallucinations, and understanding the risk they potentially pose to patient safety, ingenious mitigation strategies have recently developed. Measurement of semantic entropy [22], algorithmic approaches to address root causes of hallucinations [23], and more classical methods such as Retrieval-Augmented Generation [24] are some of the many techniques that have been proposed to identify and reduce hallucinations. Ideally, an automated combination of different strategies may help both accurately identify and reduce the occurrence of hallucinations. This would help ensure LLM response accuracy and reliability.

To remain up-to-date with these rapid and substantial developments in technology, many approaches will be required to ensure clinicians stay current and use these tools to the best of their capacity. Specific health care–related research will be required to evaluate the full extent of individual LLM capacities and performance. Furthermore, in the same way clinicians require continuing education in emerging and evolving health care topics, frequent training will be essential to use LLM-related tools adequately. Particular attention should be directed towards recognizing and mitigating the numerous pitfalls associated with their use. To this end, we have identified significant practical limitations of LLMs that may limit their rapid uptake into daily clinical practice (Textbox 2).

Textbox 2. Factors currently limiting the widespread implementation of large language models (LLMs) in clinical settings.

Model related

- Risk of bias related to:
  - The quality and quantity of the data used to train the model
  - How the model was trained and especially fine-tuned
  - How the LLM reacts to different methods of prompting
- Difficulty of complex models to explicit the reasoning behind their responses
  - Hallucination risk, without the capacity to inform the user on the final output's trustworthiness: information related to the source of the information not necessarily provided
  - No assessment of source reliability

Human related

- Risk of misuse related to:
  - Idealization of LLM capacities, and considering them to be completely foolproof, may lead to expert bias
  - Biased user input and uncritical approval of LLM output, aligning with the user's anticipated answer, can result in confirmation bias
  - Increased dependency on automated aids without critical thinking and reassessment, can lead to automation bias
- Absence of a strong legal framework defining the scope and regulatory environment of LLMs, as of date
- Lack of institutional governance defining the following:
  - Methods of informing patients and obtaining their consent for use of LLMs in their health care pathway
  - Integration of LLM use within a legal framework
  - Accountability in case of an error resulting from the use of the LLM
  - Methods of initial testing, implementation, and continuous improvement of the selected LLM
  - General operating conditions, namely:
    - Defining specific tasks for which the LLM should be used
    - Ensuring a human has the final word in the decision-making process, even though it is assisted by the LLM
  - Training procedures and certifications required for health care professionals to use the tool
  - Implementation format (on-premises vs outsourced)
  - Quality control processes
  - Traceability of feedback and changes in LLM implementation
  - Ethical framework

Economically related

- Cost of the initial investment
- Cost related to maintenance in an "on-premises" format, including:
  - Employee wages
  - Infrastructure costs
  - Energy costs
- Concerns regarding sustainability, given the high energy consumption of servers used to power and train current LLM models

As previously mentioned, future iterations of LLMs could also be specifically trained on medical datasets and fine-tuned by expert clinician input. This could be an effective method of reducing hallucinations and would allow the tailoring of LLMs to specific fields of medicine.

Conclusions

Due to its vast potential, LLM integration into routine clinical practice is no longer a question of if, but when. As technology advances, the integration of LLMs with other AI tools

possessing multimodal analysis capabilities (text, audio, and image) will follow suit. These advances offer significant opportunities in terms of patient care. However, robust legal frameworks will be necessary to guide their use on a national scale, and institutional governance is key to their implementation for everyday use. Indeed, an informed approach to using these tools, as well as significant efforts in terms of capacity-building, are primordial to avoid falling victim to their well-identified shortcomings.

Acknowledgments

No generative artificial intelligence (AI) tools were used in the creation of this manuscript.

Authors' Contributions

DR and FB contributed to writing–original draft and writing–review and editing.

Conflicts of Interest

None declared.

References

1. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. arXiv Preprint posted online on June 12, 2017 [FREE Full text] [doi: 10.48550/arXiv.1706.03762]
2. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digit Health 2023;5(6):e333-e335 [FREE Full text] [doi: 10.1016/S2589-7500(23)00083-3] [Medline: 37120418]
3. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. Radiology 2023 Jun;307(5):307(5)-e230582. [doi: 10.1148/radiol.230582] [Medline: 37191485]
4. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
5. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
6. Kung TH, Cheatham M, Medenilla A, Sillos C, de Leon L, Elepaño C, et al. Performance of chatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb 09;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
7. Dziri N, Milton S, Yu M, Zaiane O, Reddy S. On the origin of hallucinations in conversational models: is it the datasets or the models? arXiv:2204.07931 2022 Apr 17:1-14. [doi: 10.18653/v1/2022.naacl-main.387]
8. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. ACM Comput Surv 2023 Mar 03;55(12):1-38. [doi: 10.1145/3571730]
9. Azaria A, Mitchell T. The internal state of an LLM knows when its lying. 2023 Presented at: Findings of the Association for Computational Linguistics: EMNLP 2023; December 6-10, 2023; Singapore p. 967-976. [doi: 10.18653/v1/2023.findings-emnlp.68]
10. Yin Z, Sun Q, Guo Q, Wu J, Qiu X, Huang X. Do large language models know what they don't know? arXiv Preprint posted online on May 29, 2023 [FREE Full text] [doi: 10.48550/ARXIV.2305.18153]
11. Feldman P, Foulds JR, Pan S. Trapping LLM hallucinations using tagged context prompts. arXiv Preprint posted online on June 9, 2023 [FREE Full text] [doi: 10.48550/ARXIV.2306.06085]
12. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health 2023 Mar;5(3):e107-e108 [FREE Full text] [doi: 10.1016/S2589-7500(23)00021-3] [Medline: 36754724]
13. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. Semin Ophthalmol 2023;38(5):503-507. [doi: 10.1080/08820538.2023.2209166] [Medline: 37133418]
14. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023 Mar 29;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
15. Shea YF, Lee CMY, Ip WCT, Luk DWA, Wong SSW. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. JAMA Netw Open 2023;6(8):e2325000 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.25000] [Medline: 37578798]
16. Chen TC, Kaminski E, Koduri L, Singer A, Singer J, Couldwell M, et al. Chat GPT as a neuro-score calculator: analysis of a large language model's performance on various neurological exam grading scales. World Neurosurg 2023;179:e342-e347. [doi: 10.1016/j.wneu.2023.08.088] [Medline: 37634667]
17. McGowan A, Gui Y, Dobbs M, Shuster S, Cotter M, Selloni A, et al. ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. Psychiatry Res 2023 Aug;326:115334. [doi: 10.1016/j.psychres.2023.115334] [Medline: 37499282]

18. Wu RT, Dang RR. ChatGPT in head and neck scientific writing: a precautionary anecdote. Am J Otolaryngol 2023 Jul 06;44(6):103980. [doi: 10.1016/j.amjoto.2023.103980] [Medline: 37459740]
19. Medical practice variation: background paper. Australian Commission on Safety and Quality in HealthCare (ACSQHC). 2013. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/SAQ110_Medical_Practice_variation_V10_WEB.pdf [accessed 2024-04-01]
20. Chen J, Gui C, Ouyang R, Gao A, Chen S, Chen GH, et al. HuatuoGPT-vision, towards injecting medical visual knowledge into multimodal LLMs at scale. 2024 Presented at: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing; November 12-16, 2024; Miami, FL p. 7346-7370. [doi: 10.18653/v1/2024.emnlp-main.418]
21. Yin S, Fu C, Zhao S, Li K, Sun X, Xu T, et al. A survey on multimodal large language models. Natl Sci Rev 2024 Dec 23;11(12):nwae403 [FREE Full text] [doi: 10.1093/nsr/nwae403] [Medline: 39679213]
22. Farquhar S, Kossen J, Kuhn L, Gal Y. Detecting hallucinations in large language models using semantic entropy. Nature 2024 Jun 19;630(8017):625-630 [FREE Full text] [doi: 10.1038/s41586-024-07421-0] [Medline: 38898292]
23. Zhou Y, Cui C, Yoon J, Zhang L, Deng Z, Finn C, et al. Analyzing and mitigating object hallucination in large vision-language models. arXiv Preprint posted online on October 1, 2023 [FREE Full text] [doi: 10.48550/arXiv.2310.00754]
24. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv Preprint posted online on May 22, 2020 [FREE Full text] [doi: 10.48550/arXiv.2005.11401]

Abbreviations

AI: artificial intelligence
LLM: large language model

Edited by T de Azevedo Cardoso; submitted 24.04.24; peer-reviewed by J Aarts, J Luo, A Hidki; comments to author 10.09.24; revised version received 17.10.24; accepted 14.11.24; published 28.01.25.

Please cite as:
Roustan D, Bastardot F
The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations
Interact J Med Res 2025;14:e59823
URL: https://www.i-jmr.org/2025/1/e59823
doi:10.2196/59823
PMID:

Viewpoint

An Automated Clinical Laboratory Decision Support System for Test Utilization, Medical Necessity Verification, and Payment Processing

Safedin Beqaj[1], PhD; Rojeet Shrestha[2], PhD; Tim Hamill[3], MD

[1]Medical Database, Inc, Irvine, CA, United States

[2]Patient Choice Laboratories, Indianapolis, IN, United States

[3]Department of Laboratory Medicine, University of California, San Francisco, CA, United States

Corresponding Author:
Safedin Beqaj, PhD
Medical Database, Inc
1 Post, Suite 200
Irvine, CA, 92618
United States
Phone: 1 8477693701
Email: sajo@medicaldatabase.com

Abstract

Physicians could improve the efficiency of the health care system if a reliable resource were available to aid them in better understanding, selecting, and interpreting the diagnostic laboratory tests. It has been well established and widely recognized that (1) laboratory testing provides 70%-85% of the objective data that physicians use in the diagnosis and treatment of their patients; (2) orders for laboratory tests in the United States have increased, with an estimated volume of 4-5 billion tests per year; (3) there is a lack of user-friendly tools to guide physicians in their test selection and ordering; and (4) laboratory test overutilization and underutilization continue to represent a pervasive source of inefficiency in the health care system. These inappropriate test orders not only lead to slower or incorrect diagnoses for patients but also add a significant financial burden. In addition, many ordered tests are not reimbursed by Medicare because they are inappropriate for the medical condition or were ordered with the incorrect International Statistical Classification of Diseases and Related Health Problems, Tenth Revision diagnostic code, not meeting the medical necessity. Therefore, current clinical laboratory test ordering procedures experience a quality gap. Often, providers do not have access to an appropriate tool that uses evidence-based guidelines or algorithms to ensure that tests are not duplicated, overused, or underused. This viewpoint lays out the potential use of an automated laboratory clinical decision support system that helps providers order the right test for the right disease and documents the right reason or medical necessity to pay for the testing.

(Interact J Med Res 2025;14:e46007) doi:10.2196/46007

KEYWORDS

clinical decision system; CDSS; laboratory decision system; laboratory testing; test utilization; test ordering; lab test; laboratory; testing; payment; decision-making; user; utilization; processing; decision

Introduction

Laboratory testing plays a key role in clinical decision-making and physician orders for laboratory tests are increasing [1,2]. It is estimated that at least 20% of the 4-5 billion lab orders submitted annually in the United States are inappropriate. Studies have shown that overutilization and underutilization of laboratory tests occur 20.6% and 44.8% of the time, respectively [3]. This inappropriate testing not only leads to incorrect or delayed diagnoses but also significantly adds a financial burden on the health care system. This situation is expected to worsen as the available lab tests menu grows, especially in the areas of molecular diagnostics and genetic testing. Due to a lack of physician test information, education, and insurance coverage questions, ordering less effective and sometimes obsolete tests over newer tests that are more sensitive and specific remains a major problem [4]. This inappropriate testing not only led to incorrect or delayed diagnoses but also significantly added financial burden. The situation is expected to get worse as the number of lab tests is growing, especially in molecular diagnostics and genetic testing. The introduction of an automated

clinical decision support system (CDSS) that guides physicians to order the most appropriate test(s) for their patients while simultaneously providing both medical necessity requirements and applicable diagnostic codes will be a vital tool to improve test ordering and reimbursement efficiency.

Medicare and commercial health-care plans all require that ordered tests are accompanied by appropriate *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-10*) diagnostic codes that meet medical necessity rules. These requirements make it a complex process for providers to decide which tests to order, provide diagnostic information, and obtain previous authorization if required, so that test bills and payments are efficient and timely [5,6]. Given the rapidly growing demand for tests, especially molecular and genetic testing, the lack of a reliable laboratory CDSS will compound this already complex process.

When physicians fail to select and order the most appropriate test(s) based on the patient's health condition and further fail to provide the proper diagnostic codes to support medical necessity, laboratory billing will most certainly fail. The patient may then be held responsible for the laboratory charges and the laboratory will be caught in the middle of disagreements between the insurance company, the treating physician, and the patient in determining the responsible party for the laboratory charges. Ideally, every test ordered and procedure performed by the lab should be paid or reimbursed by health insurance. However, many ordered tests are not reimbursed, primarily due to a lack of medical necessity. This issue arises from ordering the wrong test that does not meet medical necessity criteria or failing to provide the correct diagnostic code for the disease or health condition. Therefore, the current clinical laboratory test-ordering procedures suffer from a quality gap and require an automated system to address this issue.

This viewpoint discusses the use of an automated laboratory CDSS that helps providers order the right test for the right disease and documents the right reason or medical necessity to pay for the testing.

Inappropriate Test Ordering

Inappropriate testing encompasses both overutilization and underutilization, both of which can affect quality patient care and health care expenditures. Overutilization includes tests that are ordered but not indicated, tests that are ordered at the incorrect time in the clinical course, or tests that are ordered too frequently. Conversely, underutilization refers to tests that are indicated but not ordered, or those that are not ordered at the appropriate time to positively impact patient care [2,5]. Both can have an adverse impact on the quality of patient care and health care costs because of downstream consequences such as additional diagnostic testing, repeat testing, imaging, prescriptions, surgeries, or prolonged hospital stays. It is estimated that more than three hundred million patients visit the laboratory annually and that at least 23 million of these patients are affected by inappropriate test ordering and test interpretation. The reports on the Commonwealth Fund Survey of Public Views of the US Health Care System (2012) found that more than 23% of laboratory tests ordered by physicians were duplicated or repeated, which increases the cost of care by further delaying or confusing the patient's diagnosis and care. It is also reported that overutilization and underutilization of laboratory tests occurred 20.6% and 44.8% of the time, respectively [2].

In 2011, a survey conducted by the Centers for Disease Control and Prevention among primary care physicians found that 14.7% had uncertainty in selecting and ordering the correct test and 8.3% had difficulty interpreting tests [7]. Physicians' lack of access to information about the availability of specific tests, the limitations of tests, and insurance coverage might play a major role in the underutilization of tests, whereas medical malpractice, selecting obsolete testing, and use of certain aspects of computerized provider order entry are documented factors leading to overutilization [3,4].

Economic Impact of Test Misutilization

Expenditures for health care in the US were approximately $4.1 trillion in 2020, which is an increase of 9.7% from 2019, and it accounts for 19.7% of the total gross domestic product. If the trends continue, health care costs are projected to increase to $6.2 trillion by 2028 [7]. Hospital and clinical service expenditure also showed rapid growth in 2020 and accounted for approximately $2.08 trillion of the total cost of health care. Although laboratory testing accounts for only a fraction of health care expenditures, 94% of objective and structured data in the electronic medical record (EMR) are obtained from a clinical laboratory [8]. Moreover, it is estimated that 60%-70% of all clinical decisions are based on the results of laboratory testing [9,10]. Considering 60% as the rate of influence on the clinical decision, it can be estimated that $1.2 trillion of health care spending is influenced by laboratory testing. Therefore, inappropriate testing not only leads to incorrect or delayed diagnoses but also significantly adds financial burden. 4-5 billion tests are performed annually in the United States. Unfortunately, it is estimated that at least 20% of the lab orders submitted are inappropriate [2]. The situation is expected to get worse as the number of esoteric lab tests is growing, especially in the areas of molecular diagnostics and genetic testing.

Ideally, every test performed by the laboratory should be reimbursed. However, many billed tests are not reimbursed due to a lack of documentation ensuring medical necessity. In many cases, the denial of the reimbursement is due to the submission of improper diagnostic code(s) for the disease or health problem being tested for. Therefore, current clinical laboratory test ordering procedures experience an information gap and there is an urgent need for an automated system to improve test utilization for economic sustainability in health care.

Need for Clinical Decision Support System in Clinical Laboratory

Selecting and obtaining authorization for appropriate medical tests is an ongoing and growing challenge in many specialties, including radiology, cardiology, pulmonology, and pharmacology. With typical radiology and diagnostic imaging costs higher than those for laboratory testing, the US government

has prioritized approval of a reimbursement reward system for insurance providers that use a CDSS to improve imaging utilization. For example, there is a 2015 "Advanced Imaging Bill" which mandates that government-approved imaging services will only be reimbursed if the insurance claim confirms that appropriate-use criteria were consulted or a CDSS was used. The bill also recommends the use of CDSS for other diagnostic test ordering, if available [5,11,12].

CDSS is currently also available in cardiology, medication management, oncology, and urology. These broad and growing applications along with expansive and expensive specialized lab testing strongly indicate that there is a substantial need for an expert laboratory CDSS to aid health care providers, care managers, and payers in selecting, ordering, and approving laboratory tests and reducing inappropriate testing.

Currently, there are some partially developed and semimanual lab CDSSs that help physicians order laboratory tests; however, these approaches are provider-driven and require inconvenient interactive user questions to access the information needed [10]. Unlike radiology CDSSs, these systems do not provide any scoring system for tests based on medical evidence, clinical relevancy, or medical necessity. Incorporating a scoring system based on test indications and providing information on supportive diagnostic codes can help automate the laboratory test ordering process and has positive impacts on test utilization, medical necessity documentation, claim verification, and payment processing. These developments strongly indicate that there is also a substantial need for a laboratory CDSS to help health care providers in selecting and ordering the appropriate laboratory tests, reduce inappropriate testing, aid providers in easier and more automated payment processing, and finally get better and on-time health care to patients [4,5,13,14].

Solution to This Problem

A potential solution is to develop a laboratory CDSS that will aid providers in selecting and ordering the right diagnostic tests with which to manage patient health care. The CDSS will help laboratories process the order, process the sample, and report accurate results on-time delivery to the ordering provider. The CDSS will provide information regarding the appropriate diagnostic *ICD-10* code(s) to meet the medical necessity. The CDSS will also provide a medical evidence-based scoring system based on clinical utility. A CDSS that provides the testing indication(s) to complement the provider's notes and is electronically interfaced with EMRs, electronic hospital records, and billing systems to automate processes like medication management and radiology CDSS is desirable [4,5,14].

Laboratory Decision System (LDS), developed by Medical Database Inc, is one of the available automated CDSS. LDS is an algorithm-based test selection and ordering database for physicians, health care providers, insurance and managed care companies, and billing services. LDS is expertly developed to help system users understand, select, order, and use laboratory tests for disease diagnosis and management using evidence-based guidelines and industry best practices. The system uses our proprietary scoring system developed by our editorial board (60 pathologists and PhDs), designed to rank testing recommendations based on disease, clinical relevance, medical necessity, and testing indication. Each time an order is placed via the LDS platform, it automatically includes the appropriate diagnostic *ICD-10* code and Local Coverage Determinations (LCD) or National Coverage Determinations (NCD) to meet medical necessity for reimbursement. Included in the robust database are all commercially available tests (over 2300 diagnostic tests), including genetic and proprietary tests [5].

The LDS rates and scores potential tests for any given disease and assigns an easily interpretable numerical and color-coded score based on clinical relevance, medical necessity, and testing indication. Tests with scores of 5 or above (10 being the highest score) meet medical necessity, while those with scores of 4 or less do not. LDS also follows Medicare's medical necessity guidelines by using testing indications such as "initial testing indication" to allow providers to better characterize the patient's disorder based on initial test results before ordering overly complex and expensive tests. Appropriate tests use indication labels, for example, diagnostic, disease management, monitoring, and alternative tests, categorizing each test with the right indication or reason for testing to avoid using providers' charts and notes that make it difficult to automate the system [4,5]. When assessing the effectiveness of LDS in improving test utilization and reimbursement with 96,170 laboratory requests comprising 374,423 test orders from a reference laboratory, 44,671 tests were accompanied by *ICD-10* that are described by Medicare as "never covered" because of the lack of a system to check or support the medical necessity of each order. A total of 160,449 tests were subject to a Medicare policy review from which 112,400 tests met coverage criteria, and 48,049 tests did not. These orders were then reevaluated using LDS. Of the original test order sample, 91.5% had an associated LDS score. Of these scored tests, 47.8% met coverage and 43.73% failed to meet coverage, according to the LDS Ranking System. Importantly, LDS provided recommendations for alternative diagnostic *ICD-10* codes or tests which could have aided physicians in choosing a more appropriate test or submitting a different *ICD-10* diagnostic code to meet medical necessity. Around 96.4% have an alternative *ICD-10* code or a test score above 5, meeting medical necessity. The LDS system recommended 80.5% which would meet Medicare policies, demonstrating that the LDS system would correct inappropriate orders if used as a testing utilization management system [5]. However, more systemic testing of the platform is needed to evaluate the effectiveness of test utilization, medical necessity verification, and payment processing.

Since the LDS platform has been built to interface with EMRs, electronic hospital records, and laboratory information systems (LIS), the content can be accessed directly through these systems, allowing orders to be sent directly to laboratories for testing coordination. Accordingly, when using the LDS platform, every test ordered will automatically include a medical necessity score, the correct testing indication, and the appropriate *ICD-10* and CPT codes, all of which also support adjudication for bill payment. An outline of the use of computerized provider order entry of LDS is elucidated in Figure 1. In addition, each test

ordered through the LDS platform will provide testing indications to support the purpose of the testing, thereby reducing manual submission of "medical necessity" review data, including reasons for test ordering (scripts, notes, charts, etc) and adding system automation with lower costs, faster throughput, and higher performance [4,15].

Figure 1. Test ordering using LDS automates the selection of appropriate tests based on clinical relevance and integrates the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision and LCD codes to facilitate reimbursement. LCD: Local Coverage Determinations; LOINC: Logical Observation Identifiers Names and Codes; NCD: National Coverage Determinations.

Conclusion

In conclusion, there is a clear and immediate need for an LDS system similar to that which is used in radiology and medication management, which can aid providers in selecting the right test for each disease or condition while assigning the correct *ICD-10* code, right Local Coverage Determinations and National Coverage Determinations to meet the medical necessity and right testing indication covering the reason and use of ordered test(s). The available LDS system developed by Medical Database, described in this viewpoint study may assist providers in making appropriate utilization decisions while also supporting laboratories in reimbursement and streamlining claim verification for payers, all of which combined will potentially make the laboratory industry and overall health care more efficient and cost-effective.

Conflicts of Interest

SB is the founder and CEO of Medical Database, Inc and developed the Laboratory Decision System.

References

1. Jackson BR. Managing laboratory test use: principles and tools. Clin Lab Med 2007;27(4):733-748. [doi: 10.1016/j.cll.2007.07.009] [Medline: 17950895]
2. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. PLoS One 2013;8(11):e78962 [FREE Full text] [doi: 10.1371/journal.pone.0078962] [Medline: 24260139]
3. Yeh DD. A clinician's perspective on laboratory utilization management. Clin Chim Acta 2014;427:145-150. [doi: 10.1016/j.cca.2013.09.023] [Medline: 24084504]
4. Hickner J, Thompson PJ, Wilkinson T, Epner P, Sheehan M, Pollock AM, et al. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. J Am Board Fam Med 2014;27(2):268-274 [FREE Full text] [doi: 10.3122/jabfm.2014.02.130104] [Medline: 24610189]
5. Leblow L, Hamill T, Beqaj SH. Use of laboratory decision system as a test utilization management tool in clinical settings, current and future perspectives. J Clin Lab Med 2019;4(1). [doi: 10.16966/2572-9578.128]
6. Sireci AN, Patel JL, Joseph L, Hiemenz MC, Rosca OC, Caughron SK, et al. Molecular pathology economics 101: an overview of molecular diagnostics coding, coverage, and reimbursement: a report of the association for molecular pathology. J Mol Diagn 2020;22(8):975-993 [FREE Full text] [doi: 10.1016/j.jmoldx.2020.05.008] [Medline: 32504675]
7. NHE fact sheet. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet [accessed 2024-12-18]
8. Forsman R. The electronic medical record: implications for the laboratory. Clin Leadersh Manag Rev 2000;14(6):292-295. [Medline: 11210217]
9. Rohr UP, Binder C, Dieterle T, Giusti F, Messina CGM, Toerien E, et al. The value of in vitro diagnostic testing in medical practice: a status report. PLoS One 2016;11(3):e0149856 [FREE Full text] [doi: 10.1371/journal.pone.0149856] [Medline: 26942417]
10. Hallworth MJ. The '70% claim': what is the evidence base? Ann Clin Biochem 2011;48(Pt 6):487-488. [doi: 10.1258/acb.2011.011177] [Medline: 22045648]
11. Bellodi E, Vagnoni E, Bonvento B, Lamma E. Economic and organizational impact of a clinical decision support system on laboratory test ordering. BMC Med Inform Decis Mak 2017;17(1):179 [FREE Full text] [doi: 10.1186/s12911-017-0574-6] [Medline: 29273037]
12. Fornell D. CMS to require appropriate use criteria documentation for medical imaging orders. Diagnostic and Interventional Cardiology. URL: https://www.dicardiology.com/article/cms-require-appropriate-use-criteria-documentation-medical-imaging-orders [accessed 2017-04-18]
13. McMorris Rodgers C. https://www.congress.gov/bill/114th-congress/house-bill/2550/text. Congress.gov. 2015 May 15. URL: https://www.congress.gov/bill/114th-congress/house-bill/2550/text [accessed 2025-02-06]
14. Baron JM, Dighe AS. The role of informatics and decision support in utilization management. Clin Chim Acta 2014;427:196-201. [doi: 10.1016/j.cca.2013.09.027] [Medline: 24084507]
15. Delvaux N, Van Thienen K, Heselmans A, de Velde SV, Ramaekers D, Aertgeerts B. The efects of computerized clinical decision support systems on laboratory test ordering: a systematic review. Arch Pathol Lab Med 2017;141(4):585-595. [doi: 10.5858/arpa.2016-0115-RA] [Medline: 28353386]

Abbreviations

CDSS: clinical decision support system
EMR: electronic medical record
ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision
LDS: Laboratory Decision System

Edited by T de Azevedo Cardoso; submitted 25.01.23; peer-reviewed by C Schäfer; comments to author 29.01.24; revised version received 13.10.24; accepted 14.01.25; published 11.02.25.

Please cite as:
Beqaj S, Shrestha R, Hamill T
An Automated Clinical Laboratory Decision Support System for Test Utilization, Medical Necessity Verification, and Payment Processing
Interact J Med Res 2025;14:e46007
URL: https://www.i-jmr.org/2025/1/e46007
doi:10.2196/46007
PMID:39808833

Viewpoint

Use of Clinical Public Databases in Hidradenitis Suppurativa Research

Xu Liu[1,2], MS; Linghong Guo[1,2*], PhD; Xian Jiang[1,2*], PhD

[1]Department of Dermatology, West China Hospital, Sichuan University, Chengdu, China

[2]Laboratory of Dermatology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Xian Jiang, PhD
Department of Dermatology
West China Hospital
Sichuan University
#37 Guoxue Alley
Wuhou District
Chengdu, 610041
China
Phone: 86 028 85423315
Fax: 86 028 85422560
Email: jiangxian@scu.edu.cn

Abstract

In this viewpoint, we argue that recent studies using clinical public databases have revolutionized our understanding of hidradenitis suppurativa (HS), a chronic inflammatory skin condition with significant impacts on patients' quality of life. Our key messages are as follows: (1) these databases enable large-scale studies integrating genetic, epidemiological, and clinical data, providing crucial insights into HS's genetic predispositions, comorbidities, and treatment outcomes; (2) findings highlight a strong genetic component, with mutations in the γ-secretase complex playing a key role in HS pathogenesis and shaping targeted therapies; (3) studies also reveal elevated risks for comorbidities like obesity, diabetes, cardiovascular disease, and systemic inflammation in patients with HS, with diet-driven inflammatory pathways potentially exacerbating disease severity; (4) while these databases offer unprecedented research opportunities, limitations such as data representativeness and quality must be considered; (5) nonetheless, their benefits outweigh potential drawbacks, allowing the identification of rare comorbidities, disease progression patterns, and personalized treatment strategies; and (6) increased funding for HS research is crucial to harness these databases' full potential, develop targeted therapies, and ultimately improve patient outcomes. As HS's impact is disproportionate to current research investments, we believe advocating for more resources and addressing database limitations will be key to advancing HS understanding and care.

(Interact J Med Res 2025;14:e70282) doi:10.2196/70282

KEYWORDS

hidradenitis suppurativa; clinical public databases; disease progression; patient data; HS

Hidradenitis suppurativa (HS) is a debilitating chronic inflammatory skin condition that significantly impacts patients' quality of life. Despite its profound effects, the pathogenesis of HS remains poorly understood [1]. In this viewpoint, we aim to highlight how recent advancements in clinical public databases have provided researchers with a powerful tool to explore the genetic, epidemiological, and clinical aspects of HS (Table 1). We argue that these databases have enabled large-scale studies integrating diverse patient data, yielding crucial insights into the disease's genetic predispositions, comorbidities, and treatment outcomes [2].

Table 1. Recent studies on hidradenitis suppurativa (HS) using clinical public databases.

Title	Year	Journal	Conclusion	Databases
Genetic Susceptibility to Hidradenitis Suppurativa and Predisposition to Cardiometabolic Disease [3]	2024	*JAMA Dermatology*	Explored the genetic susceptibility of HS to cardiovascular and metabolic diseases	UK Biobank
Association Between Hidradenitis Suppurativa and Gout: A Propensity Score-Matched Cohort Study [4]	2024	*Dermatology*	Found an association between HS and gout	TriNetX Research Network
A History of Asthma Is Associated With Susceptibility to Hidradenitis Suppurativa: A Population-Based Longitudinal Study [5]	2023	*Archives of Dermatological Research*	Found a correlation between a history of asthma and susceptibility to HS	Clalit Health Services
Hidradenitis Suppurativa and the Risk of Myocardial Infarction, Cerebrovascular Accident, and Peripheral Vascular Disease: A Population-Based Study [6]	2023	*Archives of Dermatological Research*	Investigated the risk of myocardial infarction, stroke, and peripheral vascular disease in patients with HS	Clalit Health Services
Hidradenitis Suppurativa and Rheumatoid Arthritis: Evaluating The Bidirectional Association [7]	2021	*Immunologic Research*	Evaluated the bidirectional association between HS and rheumatoid arthritis	Clalit Health Services
Association of Birth Weight, Childhood Body Mass Index, and Height With Risk of Hidradenitis Suppurativa [8]	2020	*JAMA Dermatology*	Studied the relationship between birth weight, childhood BMI, and height with the risk of HS	Danish National Patient Register
Global Hidradenitis Suppurativa COVID-19 Registry: A Registry to Inform Data-Driven Management Practices [9]	2020	*British Journal of Dermatology*	Explored data-driven management practices through the Global Hidradenitis Suppurativa COVID-19 Registry	Global Hidradenitis Suppurativa COVID-19 Registry
Comparing Cutaneous Research Funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases With 2010 Global Burden of Disease Results [10]	2014	*PLoS One*	Compared skin research funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases with the 2010 Global Burden of Disease results	National Institute of Arthritis and Musculoskeletal and Skin Diseases Database and the Global Burden of Disease

One of our key messages is that studies using clinical public databases have revealed a strong genetic component in HS. By analyzing data from diverse populations, researchers have identified genetic factors, such as mutations in the γ-secretase complex, that predispose certain individuals to develop HS [3,11]. We believe these genetic insights are not only expanding our understanding of HS pathogenesis but also shaping the development of targeted therapies [11]. Furthermore, we emphasize that clinical databases have shed light on the complex relationship between HS and various comorbidities. Studies have shown that patients with HS have a significantly higher risk of developing obesity, diabetes, cardiovascular disease, and even gout [4,6,12,13]. Notably, diet-driven inflammatory pathways, particularly those involving interleukin-17 and interleukin-1β, have been implicated in exacerbating HS severity [12]. The interplay between obesity, immune checkpoint inhibitors, and psoriasiform eruptions in HS further highlights the role of shared inflammatory pathways in disease exacerbation [13]. These findings underscore our view that a holistic approach to managing HS is needed, addressing both the dermatological and broader health risks faced by patients.

While we believe the insights gained from clinical public databases are invaluable, it is essential to consider potential limitations and counterarguments. One concern is the representativeness of the data within these databases. We argue that ensuring diverse and inclusive patient populations is crucial to avoid bias and generate findings applicable to the broader HS community [7]. Additionally, the quality and consistency of data entry across different health care settings may vary, potentially impacting the reliability of the results [5]. Furthermore, biases may exist in terms of the ethnic and geographic diversity of the data, which may skew the results for certain populations. We believe expanding the scope of these databases to include more diverse patient groups will be essential to ensure that the findings are broadly applicable and not limited by sample biases.

Despite these challenges, it is our view that the benefits of using clinical public databases in HS research far outweigh the limitations. These databases provide access to large, diverse patient populations and enable the identification of rare comorbidities and treatment outcomes that may not be apparent in smaller clinical settings. Moreover, longitudinal data captured in these databases allow researchers to study disease progression and identify early diagnostic markers, paving the way for timely interventions and personalized treatment plans [8]. We argue that researchers should focus on making improvements to database infrastructure, such as ensuring data consistency and addressing gaps in representation, to maximize their utility for future studies.

As we continue to harness the power of clinical public databases, we believe it is crucial to prioritize HS research and allocate adequate funding. The Global Burden of Disease study has

highlighted the significant impact of HS on patients' lives, yet research funding for HS remains disproportionately low compared to other skin diseases [9,10]. We argue that increased investment in HS research will accelerate the development of targeted therapies and improve patient outcomes. Public health initiatives aimed at increasing awareness of HS and its comorbidities, along with funding to support further research, will be vital to driving innovation and improving patient care.

In conclusion, it is our view that clinical public databases have revolutionized HS research by providing unprecedented access to large-scale, diverse patient data. The insights gained from these databases have deepened our understanding of the genetic susceptibility, comorbidities, and treatment outcomes associated with HS. As we move forward, we believe it is essential to address the limitations of these databases, ensure inclusive patient representation, and advocate for increased funding for HS research. By harnessing the full potential of clinical public databases, we can unlock new avenues for personalized medicine and ultimately improve the lives of individuals affected by this challenging condition.

Acknowledgments

This study was supported by the 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (ZYAI24028).

Data Availability

No new data were generated or analyzed in support of this viewpoint.

Authors' Contributions

LG (linhom.guo@foxmail.com) and XJ are cocorresponding authors.

Conflicts of Interest

None declared.

References

1. Jemec GB. Clinical practice. hidradenitis suppurativa. N Engl J Med 2012 Jan 12;366(2):158-164. [doi: 10.1056/NEJMcp1014163] [Medline: 22236226]
2. Goldburg SR, Strober BE, Payette MJ. Hidradenitis suppurativa: epidemiology, clinical presentation, and pathogenesis. J Am Acad Dermatol 2020 May;82(5):1045-1058. [doi: 10.1016/j.jaad.2019.08.090] [Medline: 31604104]
3. Nielsen VW, Bundgaard Vad O, Holgersen N, Paludan-Müller C, Meseguer Monfort L, Beyer AF, et al. Genetic susceptibility to hidradenitis suppurativa and predisposition to cardiometabolic disease. JAMA Dermatol 2025 Jan 01;161(1):22-30. [doi: 10.1001/jamadermatol.2024.3779] [Medline: 39382891]
4. Chang HC, Chiu TM, Tsai RY, Li CP, Wu YL, Chen SJ, et al. Association between hidradenitis suppurativa and gout: a propensity score-matched cohort study. Dermatology 2025 Oct 28;241(1):19-26. [doi: 10.1159/000541969] [Medline: 39467521]
5. Kridin K, Shihade W, Weinstein O, Zoller L, Onn E, Cohen A, et al. A history of asthma is associated with susceptibility to hidradenitis suppurativa: a population-based longitudinal study. Arch Dermatol Res 2023 Dec 29;315(10):2845-2851 [FREE Full text] [doi: 10.1007/s00403-023-02693-4] [Medline: 37642698]
6. Kridin K, Valido K, Cohen JM, Cohen AD. Hidradenitis suppurativa and the risk of myocardial infarction, cerebrovascular accident, and peripheral vascular disease: a population-based study. Arch Dermatol Res 2023 Apr 23;315(3):429-435. [doi: 10.1007/s00403-022-02369-5] [Medline: 35871185]
7. Kridin K, Shavit E, Damiani G, Cohen AD. Hidradenitis suppurativa and rheumatoid arthritis: evaluating the bidirectional association. Immunol Res 2021 Dec 19;69(6):533-540 [FREE Full text] [doi: 10.1007/s12026-021-09221-4] [Medline: 34409549]
8. Jørgensen AHR, Aarestrup J, Baker JL, Thomsen SF. Association of birth weight, childhood body mass index, and height with risk of hidradenitis suppurativa. JAMA Dermatol 2020 Jul 01;156(7):746-753 [FREE Full text] [doi: 10.1001/jamadermatol.2020.1047] [Medline: 32347905]
9. Naik HB, Alhusayen R, Frew J, Guilbault S, Ingram JR, Lowes MA, et al. Global Hidradenitis Suppurativa COVID-19 registry: a registry to inform data-driven management practices. Br J Dermatol 2020 Oct 02;183(4):780-781 [FREE Full text] [doi: 10.1111/bjd.19345] [Medline: 32597495]
10. Karimkhani C, Boyers LN, Margolis DJ, Naghavi M, Hay RJ, Williams HC, et al. Comparing cutaneous research funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases with 2010 Global Burden of Disease results. PLoS One 2014 Jul 8;9(7):e102122 [FREE Full text] [doi: 10.1371/journal.pone.0102122] [Medline: 25003335]
11. Chu Y, Yu S. Hidradenitis suppurativa: an understanding of genetic factors and treatment. Biomedicines 2024 Feb 01;12(2):338 [FREE Full text] [doi: 10.3390/biomedicines12020338] [Medline: 38397941]

12. Yu S, Wu X, Zhou Y, Sheng L, Jena PK, Han D, et al. A Western diet, but not a high-fat and low-sugar diet, predisposes mice to enhanced susceptibility to imiquimod-induced psoriasiform dermatitis. J Invest Dermatol 2019 Jun;139(6):1404-1407 [FREE Full text] [doi: 10.1016/j.jid.2018.12.002] [Medline: 30571973]
13. Yu S, Wu X, Shi Z, Huynh M, Jena PK, Sheng L, et al. Diet-induced obesity exacerbates imiquimod-mediated psoriasiform dermatitis in anti-PD-1 antibody-treated mice: implications for patients being treated with checkpoint inhibitors for cancer. J Dermatol Sci 2020 Mar;97(3):194-200 [FREE Full text] [doi: 10.1016/j.jdermsci.2020.01.011] [Medline: 32044178]

Abbreviations

HS: hidradenitis suppurativa

Edited by T de Azevedo Cardoso; submitted 18.12.24; peer-reviewed by X Li, S Yu; comments to author 17.01.25; revised version received 19.01.25; accepted 06.02.25; published 18.02.25.

Please cite as:
Liu X, Guo L, Jiang X
Use of Clinical Public Databases in Hidradenitis Suppurativa Research
Interact J Med Res 2025;14:e70282
URL: https://www.i-jmr.org/2025/1/e70282
doi:10.2196/70282
PMID:

Viewpoint

Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness

Liesje Donkin[1], BSc, MSc(Hons), PGDipHlthPsych, PGDipArts(Dist), PGDipClinPsych, PhD; Nathan Henry[2], BTech, MSc; Amy Kercher[1], BPsych(Hons), GDipClinPsychSci, PhD; Mangor Pedersen[1], BA, MSc, PhD; Holly Wilson[3], BSocSci(Hons), MSc; Amy Hai Yan Chan[3], BPharm(Hons), PhD

[1]Department of Psychology and Neuroscience, Auckland University of Technology, Auckland, New Zealand

[2]School of Psychological Sciences, University of Western Australia, Perth, Australia

[3]School of Pharmacy, University of Auckland, Auckland, New Zealand

Corresponding Author:
Liesje Donkin, BSc, MSc(Hons), PGDipHlthPsych, PGDipArts(Dist), PGDipClinPsych, PhD
Department of Psychology and Neuroscience
Auckland University of Technology
Private Bag 92006
Auckland, 1142
New Zealand
Phone: 64 21847886
Email: liesje.donkin@aut.ac.nz

Abstract

Internet-based research has exploded in popularity in recent years, enabling researchers to offer both investigations and interventions to broader participant populations than ever before. However, challenges associated with internet-based research have also increased—notably, difficulties verifying participant data and deliberate data manipulation by bot and spam responses. This study presents a viewpoint based on 2 case studies where internet-based research was affected by bot and spam attacks. We aim to share the learnings from these experiences with recommendations for future research practice that may reduce the likelihood or impact of future attacks. The screening and verification processes used are presented and discussed, including the limitations of these. Based on our experience, security and screening within internet-based research platforms are partly effective, but no solution is available to protect researchers completely against bot attacks. Implications for future research and advice for health researchers are discussed.

(Interact J Med Res 2025;14:e60548) doi:10.2196/60548

KEYWORDS

internet-based research; research methodology; surveys; data integrity; bot attacks; technology; data manipulation; spam; false; falsification; fraudulent; fraud; bots; research methods; data collection; verify; verification; participants

Introduction

Overview

The internet and digital technologies have irretrievably changed the conduct of research. Whether it be participant engagement, digital intervention delivery, data collection, or distribution of findings, technology has allowed researchers to engage in wider-reaching recruitment. Technology has also enabled people to access and participate in research that may not have been previously possible. In particular, health researchers have embraced internet-based recruitment to invite participation from varied population groups easily and to collect data from large samples.

There has been a significant increase in sample sizes in the last decades, intending to increase statistical power, replicability, and generalizability of research findings [1]. The adoption of open science practices and the emergence of global collaborative efforts, such as the Human Connectome Project [2] and the UK Biobank study [3], including tens of thousands of participants, has facilitated the pooling of resources and increased the overall sample sizes in research studies. However, with the increase in public accessibility of internet-based studies comes a potential increase in poor quality or false participants, ranging from

careless respondents who respond with insufficient effort [4] to malicious responses such as automated attacks in the form of computer-programmed bots [5,6].

Bots can be fully automated to target internet-based research via automated malware or malicious algorithms [7], or they may be "hybrid" where there is an element of human control. An example of a hybrid bot is when the human may complete the initial survey screening questionnaires and then allow an automated bot or algorithm to complete the remainder of the survey. Bot attacks have existed since the early 2000s and are becoming increasingly common. This increase may be linked to the rise of internet-based paid research panels and crowdsource platforms such as Amazon Mechanical Turk, which bots can leverage. Whether the bots are automated malware or human respondents who are completing surveys for financial gain without meeting study inclusion criteria [8], it is irrefutable that caution must be exercised when conducting an internet-based or remote study [9,10].

Bot attacks must be identified as they impact the integrity of datasets [5,11,12], creating a situation where ineffective interventions may appear effective, and conversely, effective interventions may seem ineffective. Similarly, one of the significant risks of undetected bot attacks is the potential to misrepresent populations [5,13] and influence decision-making based on erroneous data. Misrepresentation is particularly problematic for vulnerable [14] and at-risk populations such as Indigenous communities, who are often poorly serviced by many existing interventions [15], or populations where new effective interventions are desperately needed [13,16]. Thus, misrepresentations of bots as members of this population could have dire consequences.

This paper discusses the viewpoints of strategies researchers can use to help reduce the impact of bots on research. Like several recent papers [5,13,17-21], the researchers were impacted by bot attacks on 2 independent projects hosted on different platforms at 2 institutions. We present case studies (CS1 and CS2) of the 2 research projects (see Multimedia Appendix 1) that were affected by bot attacks and demonstrate the impact these would have had on the demographics of the final dataset. We share the steps the research teams took to identify potential bot participants, outline strategies to mitigate the bots' effects and reduce the chances of bot attacks in future projects, and provide recommendations for other researchers.

Ways to Identify Bot Responses

Several signs in both case studies indicated likely bot attacks—both in terms of researcher review of the dataset and the survey metadata and through software flagging functions. We also attempted to develop an algorithm to predict the likelihood of the participant being a bot to help determine authentic participants from bots. These strategies are discussed further.

Researcher Strategies to Identify Bots

In both studies, recruitment appeared far more efficient than expected, particularly as there was no targeted advertising and the recruitment methods used were broad. Both studies had high numbers of participant registrations in a brief period, with registrations frequently being close together (often less than minutes apart).

Second, study participant demographics were disproportionate to the sample pool, and populations often hard to recruit were overrepresented. As this was more subtle than the high number of registrations, this may not have been detected until data analysis was underway if unusual recruitment patterns were not present. Overrepresentation was particularly obvious for CS2, where we had planned to purposefully recruit based on ethnicity and monitored to ensure we were successful in this recruitment but had yet to start targeted recruitment. It is noted that although the proportion remained high in the corrected dataset, this was overinflated due to the small sample size and likely would not have remained proportionally high once higher numbers or participants were recruited.

Third, a review of participant email addresses indicated that most suspected bot attacks had emails that followed similar patterns (eg, a combination of letters and numbers followed by the exact email domains) or were nonsensical. While not all email users choose not to use their name in their emails, and the desire to protect against identity theft means that some people do not have emails that make sense, it was the similarity between the email addresses that highlighted the differences between suspected bots and the participants that we believed to be genuine.

Finally, unusual responses in the internet-based surveys also highlighted potential bot responses. Specifically, random answering for survey questions as evidenced by inconsistent or incongruent responses, a lack of answers to qualitative or text responses, and patterned survey responses indicated likely bots.

Use of Survey Platform Data

In terms of using the functionality of the survey platforms, several additional strategies can be used to determine if a bot attack has occurred. Some survey platforms can flag surveys that are likely bots, so this should be used where possible. Although responses are flagged, researchers still need to review the responses as above to identify atypical patterns of responses. While this requires researchers to review survey responses, the flagging system can potentially reduce the time required to identify bot attacks.

Potential Statistical Prediction to Identify Bots

Overview

For CS2, there were two parts to the study; participants were invited (1) to complete the "research" questionnaire, which consisted of 8 forms, and (2) to use the internet-based intervention component of the study accessed through a second external website using their contact details. By cross-validating the information from the intervention platform with the initial research survey data (N=503), we obtained confirmation of true participants (n=27). However, we could not be sure whether the high number of remaining participants were bots or actual participants who did not continue to the next part of the study, although we assumed they were bots.

Yet this scenario presented us with a unique opportunity to attempt to create a bot detection algorithm to distinguish bots from genuine participants in CS2 better using strategies previously recommended in the literature [18,22]. To achieve this, we assigned suspicion scores on a scale of 0-10 to responses that we evaluated as being potentially anomalous based on the following criteria:

Survey Completion Time

We tracked the completion rates, completion times of individual forms within the survey, and the time difference between the initial and final form completion. A weak bimodal distribution appeared for completion times in CS2, allowing us to apply a higher suspicion score to completion times of less than 5 minutes.

Email Address Analysis

We scrutinized patterns in email addresses, such as random strings of letters and digits, along with unusual domain names, to identify suspicious email patterns typically associated with bots. Higher suspicion scores were applied to unusual email address combinations.

Conflicting Response Analysis

We evaluated responses to specific correlated survey questions to detect inconsistencies that might indicate automated responses. For instance, we sought contradictory answers to 2 questions concerning the frequency of experiencing "not being able to stop or control worrying" and "worrying too much about different things" over the past 2 weeks. Greater negative correlations between paired questions were assigned higher suspicion scores.

To construct the bot detection model, we used the extreme gradient boosting (XGBoost; The XGBoost Contributors) algorithm with a logistic regression base model for binary identification of bots and participants. XGBoost is a scalable tree-boosting system that has shown excellent performance in classification tasks due to its proficiency in managing high-dimensional data and capturing intricate patterns [23,24]. We partitioned the dataset, allocating 80% for model training and 20% for testing model performance. Hyperparameters were fine-tuned through experimentation and cross-validation to attain optimal performance. We used receiver operating characteristic curve results to optimize accuracy, sensitivity, and specificity by adjusting the classification threshold. After optimization, the model achieved an overall accuracy of 0.937, with a sensitivity of 0.967 (reflecting a high rate of bot detection), but a specificity of 0.400 (reflecting a poor rate of actual participant detection), for a classification threshold of 0.25. Hence, the optimized model was heavily biased toward detecting potential bots and unable to identify true participants consistently. The model's balanced accuracy was moderately low at 0.683, reflecting the challenge of simultaneously identifying true positives and true negatives.

Based on this model, the most important features in order of predictive use were the time difference between the first and last form completions within the baseline survey, the time difference between adjacent forms from "unique" participants, the total completion time of the overall survey, the email suspicion score, and conflicting responses for correlated questions. We also ran univariate logistic regressions for individual features to validate the XGBoost model results. Heavy 0 inflation (ie, high numbers of participants with suspicion scores of 0) and poor discriminability were present in each feature, making it impossible to identify bots accurately even when features were combined.

To perform supervised learning this way, we assumed a binary classification of participants versus bots. In reality, the data were imperfectly labeled. Instead of labeling participants as "not bot" or "bot," the best we could achieve was labeling them as "not bot" or "maybe bot." Because the study comprised the internet-based research questions and the internet-based intervention, we used the login and intervention usage data to confirm some true participants (as the intervention website required interaction with diary components and activities, thus confirming a true participant). Nevertheless, there remained a high degree of uncertainty regarding the identity of the remaining participants who only completed the first baseline research questionnaire but did not proceed to log into the intervention. Thus, the bot detector's accuracy is probably exaggerated.

Ways to Manage a Bot Attack

Researchers should regularly monitor recruitment and data collection to aid in the early identification of bot attacks. Close monitoring is critical soon after studies are advertised on the web, as the risk of a bot attack, based on these 2 studies, seems to occur soon after study advertisement. While frequent data monitoring will not stop a bot attack, it may reduce the breadth of the attack by being able to intervene early. Researchers can monitor data manually by regularly checking recruitment or by setting up alerts for new enrolments, which will signal if a high volume of enrolments occurs over a short time.

Close or Pause the Survey

Once a bot attack is detected, it is essential to close or pause the survey immediately. In both case studies, new surveys were created and circulated with additional security measures in place, including location screening features, CAPTCHA coding, the use of fraud scores algorithms, and referral restrictions if recruitment occurs through social media (further strategies are discussed in the section "Ways to counteract a bot attack"). Closing or pausing the survey ensures that no new bot enrolments can occur (although settings may allow incomplete surveys to be completed) and extra security measures can be implemented.

Seek Institutional Support

After identifying the bot attack, several actions were taken to help the research teams consider what to do next. Actions included notifying the respective ethics committees about the attack and seeking consultation and advice about responding. Despite this surge in internet use for health research, there remains limited guidance from health research ethics committees on managing bot attacks or ensuring the validity of internet-based research studies. The emergence of large-scale

digital research and artificial intelligence have presented unique challenges in research ethics and safety. Artificial intelligence (AI) and "big data," particularly digital data, present unique ethical considerations, which institutional research ethics boards may not be adequately equipped to deal with [25]. For example, the issue of transparency is unique to contemporary AI models, as they are often considered "black boxes" where the user is unaware of how the model has reached its conclusion [26]. Accountability remains a complex issue as it is unclear who should be held responsible when AI is involved [27]. In this context, we believe retaining autonomy and "human control" of digital data is paramount. AI can also blur the concept of free and informed consent, and managing privacy becomes more challenging with AI's ability to identify individuals even after data deidentification [28]. Data bias is also a concern as AI may not always detect or could potentially generate biased results, including harmful gender and racial biases [29]. It is crucial to continue addressing these ethical challenges to ensure responsible and safe implementation of AI technologies.

In both our case studies, several ethical dilemmas arose. The first dilemma was the issue of reimbursement. As there is often a set budget for participant reimbursement, an ethical dilemma arises when the number of claims exceeds those budgeted. Specifically, in studies where actual enrolments exceed planned enrolments, there could potentially be hundreds of people who cannot be reimbursed if the study protocol outlines reimbursement of all respondents regardless of data quality. The second dilemma was the inability to differentiate between bots and actual participants in a way that could 100% confirm which responses were "real" and which were not. Failure to accurately identify between participants and bots could lead to a risk of reimbursing bots and erroneously excluding reimbursement of actual respondents due to budgeting constraints.

Finally, there were concerns about potential reputational risks for the researchers and their academic institutions if the research team reimbursed "fake" responses or missed payment for "real" participants. The respective university ethical and legal teams were consulted about researcher obligations to compensate based on completed answers, even if suspected bot attacks generated these. The standard advice was that the bot attack was a misrepresentative response, and therefore, there was no obligation to compensate, given that the terms of research participation had not been met.

In CS1, where participants were to be sent vouchers for compensation, careful screening of voucher claims revealed only 4 valid registrations (out of over 1200 claims), which were subsequently paid. Where participant contact details were available in CS2, we gained ethical approval to recontact participants to indicate that there had been an attack on the study. We asked participants to complete their baseline data collection again to obtain their compensation. The recompletion rate was approximately 55% and no complaints were received about this by the researchers or the ethics committee.

Remove Data

Where there was a high probability of being a bot participant, these data were removed from the dataset, and a record was made of this decision. This decision was based on the above identification parameters. Where it was difficult to tell if the participant was a bot-participant or a legitimate participant, this was noted to allow further scrutiny during the longitudinal follow-up and for reporting in publications as appropriate. It is noted that in both case studies, 2 researchers were involved in each review of data and the decision to remove participants based on being a bot. It was also determined that it was important to do this before data analysis occurred, document all research decisions, and inform the respective ethics committees.

Ways to Counteract Bot Attacks

Plan Ahead

We believe that all researchers involved in internet-based research should have a bot-management protocol in their data management plan. This could be based on protocols such as Ballard et al [11] and include the frequency of monitoring enrolments for bot attacks, steps to take when the survey is active if a bot attack is suspected, and a data handling and analysis plan for the study and that includes planning for bot-related and suspicious data. There should be a clear justification for removing data, and the dataset should be reviewed independently by 2 reviewers, including any decision made to remove data. Decisions should be documented, and publications related to the study should disclose the level of data removal related to the bot attack. Similarly, ethics committees who have approved the study should be consulted and informed about bot attacks, given the impact on research integrity.

Study Advertisements

The researchers in this study and others [9,18,30] believe that studies that are advertised on social media, particularly media where bots are highly active and reach are widespread (such as Twitter or Facebook), are more likely to experience bot attacks. The placement of study advertisements should be specific to the study population, where possible, and highlight the benefits of participating in the research other than compensation. Settings on survey platforms can also limit access to pages via routes other than where they were initially posted; for example, in CS1, settings were changed so that only participants who accessed the survey via the Facebook pages where the links were posted were included in the dataset. In addition, only those who accessed the voucher registration page via the survey could enter their details. Using previously verified distribution lists, such as those associated with professional or patient organizations, may reduce the likelihood of bot attacks as the advertising is more targeted.

Participant Information Sheets and Consent Forms

Participant information sheets and consent forms should highlight the steps researchers will take if they believe a participant has misrepresented themselves. Currently, participant information sheets often outline the inclusion and exclusion criteria and may ask that the participant indicate that they meet these criteria; however, there is no explicit statement about data management for suspected bot attacks. Upon reviewing the

participant information sheet for CS2, the ethics committee felt that the wording and context did sufficiently imply that reimbursement was given only if the questionnaires were completed as part of the intervention study, so this provided adequate grounds for researchers to omit reimbursement to participants who did not complete the questionnaires as part of the fuller study. Nevertheless, an ethics amendment was made where the wording was changed to stipulate that reimbursement more clearly was only given if the questionnaires were completed and the study intervention was used as part of participation.

For future studies, we would recommend the inclusion of text such as the following:

> *Any data that is thought to be generated by a bot or non-human means, is a duplication beyond what is required by the study, or is misrepresentative may be removed by the research team. If this occurs, study compensation for participation, as outlined in this participant information sheet, will not be offered. If you believe your data has been removed unfairly, you may discuss this with the lead researcher, who may ask you for proof of the legitimacy of your claim. If there are issues with data collection, you may be contacted to complete parts of the research again, which you may choose to decline.*

Consent forms should also include reference to this, such as "I understand that if my data is believed to be misrepresented it may be removed from the study and I may not receive compensation for participation." Before including this in external documents, we recommend consulting with your institutional legal team and ethics committee.

Prescreening of Participants

Designing a 2-step registration process that requires screening for study eligibility prior to accessing the survey and that includes multifactor authentication for identity validation may be 1 way to counter web-crawler bots. A more thorough strategy could be prescreening participants manually via a phone call, but this increases the time intensity of a study [21] and may also create a hurdle for participation by requiring additional steps to be completed. Further, commercial tools for identity verification from public records may also be used depending on the location of the study, ethics committee approval, and the researcher's ability to access these. However, these tools rely on up-to-date public records and prevent anonymous participation, which again may create barriers.

Compensation of Participants

Offering financial compensation for research participation is a common strategy when engaging in health-related research, encouraging a breadth of responses [31] and helping to reduce the impact of self-selection bias [32]. Compensating participants is considered an ethical approach to ensuring widespread health research engagement and facilitates participation by financially disadvantaged participants [33,34]. Compensation also recognizes the time participants have contributed to the research project and the value of their contribution. However, compensation means that research can also attract participants who may not be interested in the outcome of the research and are participating for other reasons.

Advertising compensation in study advertisements may increase the likelihood of studies being targeted by bots [5,6,18]. Likewise, direct compensation for internet-based data collection increases the risk of a bot attack, which may be mitigated by offering a prize draw instead. However, a prize draw may be less effective for recruitment [35], may not fairly compensate participants for their time, and may be less likely to recruit or may greater disadvantage more some individuals [36]. Alternatively, offering compensation contingent on completing specific tasks may reduce the likelihood of fraudulent participation, particularly if compensation payment is only made at the end of the study.

Ethics committees recommend offering compensation relative to the degree of involvement of the participant. For example, a 5-minute survey may only be eligible for a prize draw, whereas a study with an in-person 3-hour visit that requires blood tests and medical scans will warrant a higher compensation. However, there is limited guidance on what compensation is expected for different activities, and research teams arbitrarily determine compensation. Thus, studies where researchers offer a higher level of compensation for less effort (such as a short internet-based survey), might be more appealing to bot attacks. Given this, we would advocate for more explicit guidance on research compensation, particularly for internet-based studies.

Clarifying participant identity to claim compensation may also be another barrier for financially motivated bot attacks. Examples of this could be participants being asked to provide identification at the end of the survey to claim compensation and to ensure the participant is a person or the research teams sending vouchers to a postal address rather than an email address. Another potential way to determine authenticity may be using electronic bank transfers to a confirmed bank account rather than digital vouchers similar to case transfer programs that have been trailed in health research [37]. Alternatively, cryptocurrency could also be effective, as the process of claiming blockchain-based rewards could involve complex cryptographic tasks or multifactor authentication methods that are easy for humans but difficult for bots. Cryptocurrency transactions are also publicly verifiable by default on the blockchain, providing a transparent record of each payment (note that cryptocurrency has not yet achieved sufficiently widespread adoption to enable this method, and its use could potentially be seen as contentious [38,39]). Likewise, the requirement to share personal banking information may be a barrier to participation or could preclude people who do not have the financial fluidity to experiment with or access cryptocurrency, which has often been associated with people with more financial freedom [40]. Thus, the use of cryptocurrency may only be viable in limited situations. These strategies could reduce the potential for bots or hybrid bot attacks to be paid and could prevent participants from claiming multiple compensations for multiple completions of the same survey.

Enhanced Security Settings

Researchers should familiarize themselves with the security options within their survey platforms to reduce bot attacks. Security options include activating settings to reduce multiple responses from the same participant, bot detection features, and CAPTCHA settings. For CAPTCHA, Qualtrics (Qualtrics International Inc) generates scores for each participant on a 0 to 1 scale based on several variables. Scores less than 0.5 are likely bots. It was noted that in the case studies above, this was not enough to deter bots but was a method of helping to identify potential bots. RelevantID settings also use code to determine if participants are taking the survey multiple times, allowing researchers to remove duplicate participants.

Crafty Survey Development

Two survey design steps may help make it more difficult for bots to access surveys. The first requires registration with an email address, and then potential participants must validate their email address to log on to the internet-based survey. This registration requires more actions to log in, which makes it harder for less sophisticated bots to access the survey. However, as noted in CS2, bots or hybrid bots may be able to overcome this. Additionally, initial screening or inclusion criteria could be set up as individual questions that are randomized in order, and the correct answers need to be selected for inclusion in the final dataset. An alternative could be using questions that require 2 answers that are linked and require only certain answers to indicate a valid participant. An example could include selecting a country of origin from a drop-down list and then entering a postcode [20] which the research team check to validate if this is a real participant.

Surveys can also include questions that are only likely to be completed by bots, not participants. These are called “honey pot” questions, and each survey platform has a different way of creating them. For example, a field may be created that is invisible to humans, but can be detected by a bot. It is also noted that creating these questions in some coding software may be easier for a bot to detect, and recent internet-based threads have recommended that JavaScript (Mozilla Foundation) be the preferred coding platform for creating these fields.

The inclusion of text-based questions that need answering may also help with the identification of bot participants. Examples of this could include a series of questions requiring a typed answer about the country’s capital, what year it is, the city the person lives in, and basic mathematical equations such as “What is one plus three?” Question format can also increase the chance of detecting a bot by directing participants only to select a specific number of items from a list where all items could be answered.

Kay and Saucier [41] have created an internet-based database of 660 items that can be used to detect careless or insufficient-effort responders in survey data. The authors propose that the best way to identify participants who respond in a misleading or invalid way is to add frequency and infrequency items. Frequency items should be endorsed almost by every respondent (eg, it should be illegal to kill an innocent person), and infrequency item statements should be endorsed by almost no respondents (eg, when someone tells a funny joke, I feel angry) [41]. Thus, failure to endorse a frequency item or an infrequency item would likely indicate a bot.

The final recommendation is randomization of survey items or questions that make it difficult for bots to “predict” the question and answer “appropriately,” thus making it easier to identify nonsensical surveys. Similarly, including free text or open text questions may also help to identify bots versus actual participants, as bots are less likely to complete these meaningfully or may replicate answers. Thus, reviewing qualitative answers for quality or content duplication may help identify a bot. Additionally, including qualitative answers may make bot responding more challenging and provide more insights for future research not captured in closed-text answers [14]. It is noted that these strategies individually will not necessarily prevent hybrid people-bot attacks but will increase the difficulty of survey completion. If multiple strategies are used, it will likely be more deterring for the bot and make it easier to identify bot-responding.

Metadata

The use of IP address tracking may be helpful to identify actual participants compared to bots. Coupled with location filters, IP address tracking can help determine when multiple attempts are made of the same survey and when the participant is not within the geographical catchment of the survey. However, it is noted that virtual private networks can circumnavigate this, and IP address tracking alone is not foolproof. Some ethics committees consider IP addresses to be identifiable information that must be disclosed during the informed consent process if collected. Using this strategy should be approved by the overseeing ethics committee. For further details on using IP addresses to mitigate bots’ effects, see White and Brodhead [8].

Asking participants to provide information that is matched to their IP address or metadata could also be another strategy to validate real participants. This could include strategies such as asking what country the person is in and matching it to an IP address or geolocation or to confirm the time where they are and match it to survey functions which capture the times of both the participant’s browser and the server of the host institution to detect discrepancies [20].

Challenges

Overview

There is likely no foolproof way to detect bots and determined bot users could overcome many of these strategies through a human answering some of the screening questions and then running the bot script. Thus, bot-attacks in internet-based research are an ongoing problem to manage. There are also several ethical issues when identifying potential bots.

Decreased Anonymity

Requiring participants to provide a form of identification prevents researchers from collecting anonymous responses. This may be particularly problematic for studies where participants may be reluctant to participate due to fear, stigma, or where the behavior being studied is illegal [42]. There are strategies to

reduce the impact of the potential to be identified, such as the software randomly allocating codes that are not tied to answers or a researcher who does not have access to the survey data (or even the topic) being the contact person for voucher claims. Researchers working in areas where stigma and shame are common may be reluctant to design surveys where participants may need to be identified; however, there is research showing that larger incentives and greater privacy do not necessarily equate to disclosure of more sensitive data [43] and greater anonymity may impact the accuracy of internet-based surveys [44]. Thus, protecting participant identity through deidentification may be the preferred route to prevent bots and increase survey accuracy.

Removal of Data

Removing bot data is important for research integrity, but this should be done cautiously so as not to remove data belonging to valid participants. This is particularly valid when considering removing data with incomplete answers (particularly qualitative answers), as this may result in removing data from people who may struggle with literacy but also misrepresent attrition and engagement—both critical issues in digital interventions [45-47].

Data removal should be done by 2 independent, blinded researchers and disclosed in publications to ensure transparency and rigor and decrease biases. Similarly, sharing the processes around managing bot-suspected material will aid the research community in developing a consistent approach to bot-data management. Specifically, we believe that researchers should report the methods used to determine bot data, the percentage of data removed due to the data being bot-generated, and the percentage of any suspicious remaining data. As such, we need consistent community guidelines for handling and reporting bot-related data and research impact.

Abusive Responses or Researcher Protection

Soon after the early closure of one of the studies, the researchers began to receive requests for participation compensation. When "participants" were told that there would be no compensation due to suspected fraudulent behavior, researchers began to receive abusive messages. While this is often not at the level of abuse that has recently been disclosed with the rise of the internet [48,49], receiving such abuse is a risk to the well-being of researchers [50,51]. Given this, a plan is needed to determine if and how to respond to such communications and ensure the researcher's well-being [50,51].

Discussion

Principal Findings

This study presents 2 case studies of bot attacks encountered by our research team and the strategies we used to counter them. Moreover, it underscores the critical challenge facing internet-based survey research, which remains highly susceptible to various forms of malicious interference. Unless novel technologies are developed to counter these threats, the challenge of obtaining trustworthy survey data over the internet will likely become more difficult in future years. Our hope is that this paper can serve as a warning, helping to drive the changes needed to protect the integrity of internet-based survey research moving forward.

As technology evolves, so too will the sophistication of bot attacks. As such, bots are here to stay. Therefore, researchers will need multiple, regularly updated strategies to combat and manage bots' effects, including institutional support and ethics committees knowledgeable in this space. It is likely that health research will need to consider multiple bot prevention strategies, use multimodal recruitment and data collection, and develop clear guidelines to help researchers manage bot-related data in internet-based research.

Authentic participant data are valuable, and therefore, researchers need algorithms and a decisional process to detect bot data. A method that is overly sensitive to bots may remove actual participants unnecessarily, thus reducing the size of the dataset and may result in the study being underpowered. On the other hand, an algorithm or decisional process with low specificity may allow an excessive number of bot data to be included, reducing data quality and increasing study costs through compensation paid to nonauthentic participants. Further, bots are likely to provide either extremely noisy or biased answers that negatively impact data quality. When creating a bot detection algorithm, it is up to the researcher to decide whether they wish to prioritize sensitivity (the ability to detect participants) or specificity (the ability to detect bots).

When attempting to develop an appropriate algorithm to detect bots, we faced severe issues with zero-inflated data where most participants—whether bots or not—registered a suspicion score of 0 for multiple variables. This indicates a lack of information about the methods used to create the bot. It highlights the possibility that the attack was performed by human actors responding manually to each survey, making it seem more authentic. Yet even with perfect knowledge of our artificial bot dataset, any bot detection algorithm trained on this dataset would endure from severe overfitting issues, as the attack strategies used will likely differ from those used for another project. Consequently, our detection algorithm probably would not generalize well to real-world attacks, except for similarly designed REDCap (Research Electronic Data Capture; Vanderbilt University) surveys. To make this even more challenging, each of the survey tools currently available to researchers (such as Qualtrics, REDCap, or SurveyMonkey) have different programming interfaces and database structures that would each require a unique version of the bot detection algorithm that is configured to their setup. Hence, we remain unaware of any tool that provides sufficient accuracy for bot detection across the wide range of currently used survey tools. Detection is made more difficult by imperfect participant labeling. For example, researchers can rarely be completely certain that an email address belongs to a bot, even in the presence of "obvious" signs such as random strings of letters or digits. Genuine participants wishing to preserve their anonymity on the web may create a dummy email address for responding to surveys or use "hide my email address" systems, which can also create unusual email handles. The best way to counter fake email addresses is by screening potential participants beforehand and sending a private, individual survey link to prescreened addresses rather than making the link

publicly available [18]. However, this may not be practical for all projects as it can be labor-intensive and as highlighted in CS2, is not foolproof.

Implications

Health research often engages vulnerable populations, and special care must be taken to protect privacy and ensure appropriate risk management is available where needed—including the detection of bots that may skew findings. One vulnerable group are the Indigenous populations. Health research must be responsive and capture the needs of indigenous populations who often experience poorer health outcomes. In CS1, bot responses identified as Māori (the Indigenous population of New Zealand) made up 30% of the sample, and after bot measures were in place, the percentage identifying as Māori was 10%. In CS2, "bot" responses inflated the participants that identified as Māori to almost 4 times what we expected in recruitment based on the population composition. Had we not identified the bot, we would have been quite confident in the analysis we made relating to Māori participants, given the high number, despite these results being unrelated to Māori at all. Therefore, there is a risk that failure to detect and adapt to bot technology could lead to misinformation and contribute to poorer outcomes for indigenous populations based on misinformed conclusions.

Given this, we encourage researchers to consider whether internet-based surveys are the best way to obtain a representative sample with high data accuracy. Internet-based recruitment may preclude some people, and although we increasingly see people being connected on the web, regular data connection, access to technology, and technology literacy may not yet be everyone's privilege, including vulnerable communities. This is particularly pertinent for those where the digital world may be at odds with core and cultural values [52] or where surveys may not accurately capture experiences [53]. Instead, the use of multimodal surveys may reach a broader group, including those who may not be aware of internet-based surveys, may be more responsive to the needs of communities, and may be able better to manage the impact of bots [14].

Recommendations

From our experiences and literature review, it would seem prudent to develop clear guidelines for conducting internet-based research to reduce the risk of bot attacks and increase the robustness of internet-based research. There are currently few guidelines that exist to guide practice in this area. The Association of Internet Researchers released their latest ethical guidelines for internet research in 2019 [54]. The guidelines provide useful information about data management and security, and consider some of the critical issues that we discuss in our paper, including how to protect the researcher where the researcher's public identity is known; specific ethical topics such as accountability, trust, and transparency which have different considerations for internet-based research; and issues related to the accuracy of data including in-built biases from algorithms used for collection and how to use metadata. The EQUATOR (Enhancing the Quality and Transparency of Health Research) network refers to 3 reporting guidelines relating to digital health research—the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist—a checklist for reporting results of internet e-surveys [55]; the CONSORT (Consolidated Standards of Reporting Trials)-EHEALTH to standardize the reporting of evaluations of web and mobile interventions [56], and the iCHECK-DH (Guidelines and Checklist for the Reporting on Digital Health Implementations) which is a guideline and checklist on how to report digital health implementations [57,58]. However, these EQUATOR guidelines refer more to the reporting of the internet-based study than the conduct and do not refer to bot attacks or management [59]. We advocate for developing standardized guidance for researchers conducting internet-based research that describe key considerations and a standardized approach to ensuring data accuracy and validity. These guidelines should also include recommendations for reporting when bot attacks may have occurred and how researchers should handle data with regard to this. While we recognize that bot attacks are likely to be ever-evolving and are likely to outpace the development of guidelines, we believe it would still be useful for guidelines or discussion documents for researchers for strategies to mitigate or minimize the potential effects of bots. Similarly, these guidelines or discussion documents may include points to help researchers determine if the strategies required to minimize bot attacks and the probability of obtaining valid data outweighs the risk of an attack, thus determining if internet-based health research is viable.

Future Directions

Bot attacks will likely become more widespread and more difficult for researchers to detect as botware and attack algorithms become more sophisticated. The experiences of our research group and others [5,13,17-19] highlight a critical, growing problem that deserves more focused researcher attention. Nevertheless, internet-based research has many advantages, such as the ability to reach large numbers of participants and the ability to complete the research remotely. Our experiences should not deter researchers from conducting internet-based research studies. Instead, our paper is a call to action to raise awareness and encourage researchers to consider the risks and benefits of internet-based research. We recommend developing guidelines around detecting and managing bot data in internet-based surveys to help raise awareness of these issues, provide guidance around survey design and data management, and encourage transparency in reporting data that bots may have impacted.

Researchers should also consider the recent exponential development of large language models (LLMs) such as ChatGPT (OpenAI) [60]. Unlike manually programmed bots, LLM-assisted bots can interpret and respond to surveys more coherently, making LLM responses more challenging to detect. This also allows attackers to automate responses to qualitative questions, a task previously reliant on human guidance. We believe it is only a matter of time before LLM-assisted bots become sophisticated enough to respond to any survey design. Consequently, it may become impossible to trust any results obtained from public survey links. Therefore, we recommend that researchers begin implementing more effective security strategies if they have not already done so.

A question that arises at this point, and a potential focus for further research, is whether bot attacks are generalizable across different parts of the world, as different countries have varying protocols for paying research participants [61]. Our experience also suggests that all current research platforms may be vulnerable to illegal bot attacks, and the associated responsibility of software developers to ensure the security and privacy of people is paramount. Thus, bot attacks are likely to be a global issue.

Conclusions

While internet-based research studies increase the ease of participant recruitment and accessibility to a diverse range of respondents, the rise of sophisticated bot programmers and algorithms to automate survey responses risks invalidating internet-based research. Careful planning of internet-based research study designs and incorporating measures to minimize bot responses, such as using a mix of closed, open-ended, and randomized questions, is necessary to protect internet-based studies from bot attacks. However, these measures should be weighed against the risk of inadvertently disqualifying or turning away real, genuine participants. There is an urgent need for standardized practices and guidelines to be developed to provide researchers with clear guidance on safeguarding against bot attacks and actions to take if a bot attack is suspected. As bot attacks are here to stay, this paper aims to raise researchers' awareness and create a call to action before the problem becomes more widespread and challenging to manage.

Authors' Contributions

LD, NH, MP, AK, and AHYC performed the conceptualization. LD, NH, and HW contributed to the data curation. NH, HW, and LD did the formal analysis. LD, NH, AK, and AHYC did the methodology. LD, NH, AK, and AHYC contributed to the investigation. NH contributed to the software. NH and MP contributed to the validation. LD, NH, AK, HW, and AHYC contributed to writing the original draft. LD, NH, AK, MP, HW, and AHYC contributed to writing—reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Case studies.
[DOCX File , 34 KB - ijmr_v14i1e60548_app1.docx]

References

1. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol 2019;2(1):9 [FREE Full text] [doi: 10.1038/s42003-018-0261-x] [Medline: 30623105]
2. van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: an overview. Neuroimage 2013;80:62-79 [FREE Full text] [doi: 10.1016/j.neuroimage.2013.05.041] [Medline: 23684880]
3. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779 [FREE Full text] [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
4. Ward M, Meade AW. Dealing with careless responding in survey data: prevention, identification, and recommended best practices. Annu Rev Psychol 2023;74(1):577-596 [FREE Full text] [doi: 10.1146/annurev-psych-040422-045007] [Medline: 35973734]
5. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022;56(4):2841-2852 [FREE Full text] [doi: 10.1007/s11135-021-01252-1] [Medline: 34629553]
6. Lawlor J, Thomas C, Guhin AT, Kenyon K, Lerner MD, Drahota A. Suspicious and fraudulent online survey participation: introducing the REAL framework. Methodological Innovations 2021;14(3). [doi: 10.1177/20597991211050467]
7. Dittrich D, Leder F, Werner T. A Case Study in Ethical Decision Making Regarding Remote Mitigation of Botnets. Berlin, Germany: Springer; 2010.
8. White AN, Brodhead MT. Detecting fraudulent responses in online survey research. PsyArXiv Preprints Preprint posted online on May 26, 2023 [FREE Full text] [doi: 10.31234/osf.io/wuk2y]
9. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
10. Smyk M, Tyrowicz J, van der Velde L. A cautionary note on the reliability of the online survey data: the case of wage indicator. Sociol Methods Res 2018;50(1):429-464. [doi: 10.1177/0049124118782538]
11. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. JMIR Public Health Surveill 2019;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]

12. Dupuis M, Meier E, Cuneo F. Detecting computer-generated random responding in questionnaire-based data: a comparison of seven indices. Behav Res Methods 2019;51(5):2228-2237. [doi: 10.3758/s13428-018-1103-y] [Medline: 30091086]
13. Bybee S, Cloyes K, Ellington L, Baucom B, Supiano K, Mooney K. Bots and nots: safeguarding online survey research with underrepresented and diverse populations. Psychol Sex 2022;13(4):901-911 [FREE Full text] [doi: 10.1080/19419899.2021.1936617] [Medline: 36439051]
14. Yazici E, Wang Y. Attack the bot: Mode effects and the challenges of conducting a mixed-mode household survey during the Covid-19 pandemic. Int J Soc Res Methodol 2023;27(6):791-796. [doi: 10.1080/13645579.2023.2241797]
15. Barrett NM, Tamatea J, Morgan R, Jones AR. Creating an environment to inform, build, and sustain a Maori health research workforce. J R Soc N Z 2023;55(1):47-61. [doi: 10.1080/03036758.2023.2235303]
16. Corsini-Munt S, Rancourt KM, Dubé J, Rossi M, Rosen N. Vulvodynia: a consideration of clinical and methodological research challenges and recommended solutions. JPR 2017;10:2425-2436. [doi: 10.2147/jpr.s126259]
17. Loebenberg G, Oldham M, Brown J, Dinu L, Michie S, Field M, et al. Bot or not? Detecting and managing participant deception when conducting digital research remotely: case study of a randomized controlled trial. J Med Internet Res 2023;25:e46523 [FREE Full text] [doi: 10.2196/46523] [Medline: 37707943]
18. Storozuk A, Ashley M, Delage V, Maloney EA. Got bots? Practical recommendations to protect online survey data from bot attacks. TQMP 2020;16(5):472-481. [doi: 10.20982/tqmp.16.5.p472]
19. Goodrich B, Fenton M, Penn J, Bovay J, Mountain T. Battling bots: experiences and strategies to mitigate fraudulent responses in online surveys. Appl Econ Perspect Pol 2023;45(2):762-784. [doi: 10.1002/aepp.13353]
20. Wang J, Calderon G, Hager ER, Edwards LV, Berry AA, Liu Y, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. PLOS Glob Public Health 2023;3(8):e0001452 [FREE Full text] [doi: 10.1371/journal.pgph.0001452] [Medline: 37610999]
21. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! identifying eligibility fraud by applicants in digital health research. Internet Interv 2021;25:100401 [FREE Full text] [doi: 10.1016/j.invent.2021.100401] [Medline: 34094883]
22. Xu Y, Pace S, Kim J, Iachini A, King LB, Harrison T, et al. Threats to online surveys: recognizing, detecting, and preventing survey bots. Soc Work Res 2022;46(4):343-350 [FREE Full text]
23. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 2020;54(3):1937-1967. [doi: 10.1007/s10462-020-09896-5]
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
25. Bouhouita-Guermech S, Gogognon P, Bélisle-Pipon JC. Specific challenges posed by artificial intelligence in research ethics. Front Artif Intell 2023;6:1149082 [FREE Full text] [doi: 10.3389/frai.2023.1149082] [Medline: 37483869]
26. Brożek B, Furman M, Jakubiec M, Kucharzyk B. The black box problem revisited. real and imaginary challenges for automated legal decision making. Artif Intell Law 2023;32(2):427-440. [doi: 10.1007/s10506-023-09356-9]
27. Sand M, Durán JM, Jongsma KR. Responsibility beyond design: physicians' requirements for ethical medical AI. Bioethics 2022;36(2):162-169 [FREE Full text] [doi: 10.1111/bioe.12887] [Medline: 34089625]
28. Erlich Y, Shor T, Pe'er I, Carmi S. Identity inference of genomic data using long-range familial searches. Science 2018;362(6415):690-694 [FREE Full text] [doi: 10.1126/science.aau4832] [Medline: 30309907]
29. Jain A, Brooks JR, Alford CC, Chang CS, Mueller NM, Umscheid CA, et al. Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms. JAMA Health Forum 2023;4(6):e231197 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.1197] [Medline: 37266959]
30. Orabi M, Mouheb D, Al Aghbari Z, Kamel I. Detection of bots in social media: a systematic review. Inf Process Manag 2020;57(4):102250. [doi: 10.1016/j.ipm.2020.102250]
31. Abdelazeem B, Abbas KS, Amin MA, El-Shahat NA, Malik B, Kalantary A, et al. The effectiveness of incentives for research participation: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2022;17(4):e0267534 [FREE Full text] [doi: 10.1371/journal.pone.0267534] [Medline: 35452488]
32. Kaźmierczak I, Zajenkowska A, Rogoza R, Jonason PK, Ścigała D. Self-selection biases in psychological studies: personality and affective disorders are prevalent among participants. PLoS One 2023;18(3):e0281046 [FREE Full text] [doi: 10.1371/journal.pone.0281046] [Medline: 36888578]
33. Jennings CG, MacDonald TM, Wei L, Brown MJ, McConnachie L, Mackenzie IS. Does offering an incentive payment improve recruitment to clinical trials and increase the proportion of socially deprived and elderly participants? Trials 2015;16(1):80 [FREE Full text] [doi: 10.1186/s13063-015-0582-8] [Medline: 25888477]
34. Permuth-Wey J, Borenstein AR. Financial remuneration for clinical and behavioral research participation: ethical and practical considerations. Ann Epidemiol 2009;19(4):280-285. [doi: 10.1016/j.annepidem.2009.01.004] [Medline: 19230712]
35. Dykema J, Stevenson J, Jones CP, Day B. Guaranteed incentives and prize drawings: effects on participation, data quality, and costs in a web survey of college students on sensitive topics. Soc Sci Comput Rev 2023;42(3):777-793. [doi: 10.1177/08944393231189853]

36. Bierer BE, White SA, Gelinas L, Strauss DH. Fair payment and just benefits to enhance diversity in clinical research. J Clin Transl Sci 2021;5(1):e159 [FREE Full text] [doi: 10.1017/cts.2021.816] [Medline: 34527298]
37. Lagarde M, Haines A, Palmer N. Conditional cash transfers for improving uptake of health interventions in low- and middle-income countries: a systematic review. JAMA 2007;298(16):1900-1910. [doi: 10.1001/jama.298.16.1900] [Medline: 17954541]
38. Valdeolmillos D, Mezquita Y, González-Briones A, Prieto J, Corchado JM. Blockchain Technology: A Review of the Current Challenges of Cryptocurrency. Cham, Switzerland: Springer International Publishing; 2020.
39. Alzahrani S, Daim TU. Analysis of the cryptocurrency adoption decision: literature review. 2019 Presented at: Portland International Conference on Management of Engineering and Technology (PICMET); August 25-29, 2019; Portland, OR. [doi: 10.23919/picmet.2019.8893819]
40. Steinmetz F, von Meduna M, Ante L, Fiedler I. Ownership, uses and perceptions of cryptocurrency: results from a population survey. Technol Forecast Soc Change 2021;173:121073. [doi: 10.1016/j.techfore.2021.121073]
41. Kay CS, Saucier G. The comprehensive infrequency/frequency item repository (CIFR): an online database of items for detecting careless/insufficient-effort responders in survey data. Personal Individual Differ 2023;205:112073. [doi: 10.1016/j.paid.2022.112073]
42. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant. Int J Qual Methods 2015;14(5). [doi: 10.1177/1609406915621420]
43. Murdoch M, Simon AB, Polusny MA, Bangerter AK, Grill JP, Noorbaloochi S, et al. Impact of different privacy conditions and incentives on survey response rate, participant representativeness, and disclosure of sensitive information: a randomized controlled trial. BMC Med Res Methodol 2014;14(1):90 [FREE Full text] [doi: 10.1186/1471-2288-14-90] [Medline: 25027174]
44. Lelkes Y, Krosnick JA, Marx DM, Judd CM, Park B. Complete anonymity compromises the accuracy of self-reports. J Exp Soc Psychol 2012;48(6):1291-1299. [doi: 10.1016/j.jesp.2012.07.002]
45. Borghouts J, Eikey E, Mark G, de Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
46. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
47. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
48. Bisaillon J, Mercure C, Villeneuve S, Plante I. Cyberbullying of professors: what measures are in place in universities and what solutions are proposed by victims? Stud Higher Educ 2023;48(11):1639-1650. [doi: 10.1080/03075079.2023.2206431]
49. Doerfler P, Forte A, de Cristofaro E, Stringhini G, Blackburn J, McCoy D. "I'm a professor, which isn't usually a dangerous job": internet-facilitated harassment and its impact on researchers. Proceedings of the ACM on Human-Computer Interaction 2021 Oct 18;5(CSCW2):1-32. [doi: 10.1145/3476082]
50. Gelms B. Social media research and the methodological problem of harassment: foregrounding researcher safety. Comput Compos 2021;59:102626. [doi: 10.1016/j.compcom.2021.102626]
51. Mattheis AA, Kingdon A. Does the institution have a plan for that? Researcher safety and the ethics of institutional responsibility. In: Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches. Cham, Switzerland: Palgrave Macmillan; 2021:457-472.
52. Donkin L, Bidois-Putt M, Wilson H, Hayward P, Chan AHY. An exploration of the goodness of fit of web-based tools for māori: qualitative study using interviews and focus groups. JMIR Form Res 2024;8:e50385 [FREE Full text] [doi: 10.2196/50385] [Medline: 38696236]
53. Holcombe-James I. Survey-based research in remote Indigenous communities: considerations for methods. In: Rennie E, editor. Field Guide to Intercultural Research. Cheltenham, United Kingdom: Edward Elgar Publishing; 2021.
54. Heise AHH. Internet Research: Ethical Guidelines 3.0. WildApricot. 2019. URL: https://aoir.org/reports/ethics3.pdf [accessed 2019-10-04]
55. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
56. Eysenbach GE. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
57. Perrin Franck C, Babington-Ashaye A, Dietrich D, Bediang G, Veltsos P, Gupta PP, et al. iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations. J Med Internet Res 2023;25:e46694 [FREE Full text] [doi: 10.2196/46694] [Medline: 37163336]
58. Perrin Franck C, Babington-Ashaye A, Dietrich D, Bediang G, Veltsos P, Gupta PP, et al. Correction: iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations. J Med Internet Res 2023;25:e49027 [FREE Full text] [doi: 10.2196/49027] [Medline: 37201181]

59. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. Lancet 2008;371(9619):1149-1150. [doi: 10.1016/s0140-6736(08)60505-x]
60. Kumar S, Musharaf D, Musharaf S, Sagar AK. A Comprehensive Review of the Latest Advancements in Large Generative AI Models. Cham, Switzerland: Springer Nature; 2023.
61. Grady C, Dickert N, Jawetz T, Gensler G, Emanuel E. An analysis of U.S. practices of paying research participants. Contemp Clin Trials 2005;26(3):365-375. [doi: 10.1016/j.cct.2005.02.003] [Medline: 15911470]

Abbreviations

AI: artificial intelligence
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT: Consolidated Standards of Reporting Trials
CS: case study
EQUATOR: Enhancing the Quality and Transparency of Health Research
iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations
LLM: large language model
REDCap: Research Electronic Data Capture
XGBoost: extreme gradient boosting

Edited by T de Azevedo Cardoso; submitted 15.05.24; peer-reviewed by E Frechman, L Johnson; comments to author 07.08.24; revised version received 29.08.24; accepted 24.09.24; published 14.03.25.

Please cite as:
Donkin L, Henry N, Kercher A, Pedersen M, Wilson H, Chan AHY
Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness
Interact J Med Res 2025;14:e60548
URL: https://www.i-jmr.org/2025/1/e60548
doi:10.2196/60548
PMID:

Review

The Effect of Combining mHealth and Health Professional–Led Intervention for Improving Health-Related Outcomes in Chronic Diseases: Systematic Review and Meta-Analysis

Masashi Kanai[1,2*], PhD; Takahiro Miki[1*], PhD; Toshiya Sakoda[1], MSc; Yuta Hagiwara[1], MSc

[1]PREVENT Inc, Aichi, Japan

[2]Institute of Transdisciplinary Science for Innovation, Kanawaza University, Kanazawa, Japan

[*]these authors contributed equally

Corresponding Author:
Takahiro Miki, PhD
PREVENT Inc
IKKO Shinsakae Building 9F
1 Aoi, Higashi-ku
Aichi, 461-0004
Japan
Phone: 81 527157955
Email: miki.takahiro@prevent.co.jp

Abstract

Background: Chronic diseases such as diabetes and cardiovascular disease are global health challenges, affecting millions of people worldwide. Traditional health care often falls short in chronic disease management. This has led to the exploration of innovative solutions, such as mobile health (mHealth) technologies. mHealth, which leverages mobile and wireless technologies, has the potential to transform health care delivery by providing continuous, accessible, and personalized care. However, the effectiveness of mHealth, particularly when integrated with traditional health care interventions delivered by professionals, warrants comprehensive investigation. Understanding the combined impact of mHealth and professional-led interventions is critical to maximizing the potential of mHealth to improve patient outcomes and adherence.

Objective: This study aims to investigate the effectiveness of combining mHealth and health professional–led intervention for improving health-related outcomes in chronic diseases

Methods: This systematic review and meta-analysis focused on randomized controlled trials. We searched Web of Science, CENTRAL, MEDLINE, and CINAHL through July 17, 2023. The study targeted patients aged 18 years and older, experiencing at least 1 chronic condition. The interventions were a combination of mHealth and the use of a health care professional. The comparison groups consisted of participants receiving either general care and follow-up or those using mHealth devices without any health care professional involvement. The outcomes measured in this review included hemoglobin A_{1c} (HbA_{1c}), quality of life (QoL), and physical activity.

Results: The study included 26 research papers, encompassing 7360 individuals. Meta-analysis was conducted for HbA_{1c}, QoL, and physical activity. For HbA_{1c}, short-term improvement was significant (standardized mean difference [SMD] –0.43; 95% CI –0.64 to –0.21; I^2=69%) and medium term (SMD –0.49; 95% CI –0.49 to –0.09; I^2=21%). However, in the long term, the improvement was not significant (SMD –0.23; 95% CI –0.49 to 0.03; I^2=88%). For QoL, significant improvements were observed in the short term (SMD –0.23; 95% CI –0.42 to –0.05; I^2=62%), and in the medium term (SMD –0.16; 95% CI –0.24 to –0.07; I^2=0%). In the long term, however, the improvement was not significant (SMD –0.12; 95% CI –0.41 to 0.16; I^2=71%). For physical activity, both subjective (questionnaire) and objective (number of steps) outcomes were analyzed. In the short term, subjective outcomes showed significant improvement (SMD 0.31; 95% CI 0.12-0.50; I^2=0%), while objective outcomes did not (SMD 0.11; 95% CI –0.05 to 0.27; I^2=0%). Medium- and long-term subjective outcomes showed no significant improvement. Meta-analysis for objective outcomes in the medium and long term was not possible due to insufficient studies.

Conclusions: This study confirmed short- and medium-term benefits of mHealth combined with professional interventions for HbA_{1c}, QoL, and short-term physical activity, supporting effective chronic disease management.

(Interact J Med Res 2025;14:e55835) doi:10.2196/55835

KEYWORDS

mHealth; systematic reviews; meta-analysis; chronic diseases; global health; technology; health care; interventions; chronic conditions; health care professionals; World Health Organization; physical activity; web-based

Introduction

Background

Mobile health (mHealth) is the delivery of medical and public health services using mobile handsets, mobile kiosks, and other wireless terminals in real-world environments [1]. Interventions in mHealth can take various forms, including the creation of original apps for the management of health conditions and communication, simple text message-based interventions, as well as those using voice calls or the accumulation of device data on mobile apps for data management [2]. Many interventions in the field of mHealth primarily emphasize self-management without any direct human involvement [3]. Likewise, there are interventions that involve periodic health care professional assistance and the additional use of health devices [4,5].

The growing disease burden of chronic diseases, on the other hand, is one of the global challenges [6]. Chronic diseases include conditions such as cardiovascular diseases, stroke, cancer, diabetes, respiratory conditions, and arthritis [7]. Furthermore, obesity is now defined as a disease, and the most common chronic disease [8]. While the assistance of health care professionals is essential for the management of these chronic conditions, mHealth may be valuable in filling the gaps between medical consultations or support, as well as in the monitoring of disease conditions remotely. In fact, the efficacy of therapeutic apps prescribed by physicians has also been reported [9]. The effectiveness of mHealth is currently being investigated through a systematic review [10]. The quality of research methodologies is often limited, and consistent conclusions have not been reached yet. Among the reasons for this is that the definition of mHealth intervention methods varies widely, and depending on the definition, conclusions about effectiveness can differ. Some reports suggest that no significant effects have been observed. There is a discrepancy in research on health interventions that focus on self-monitoring for chronic diseases, and it is unclear if there are any conclusive benefits [11-13]. A systematic review evaluating the effectiveness of remote monitoring in various types of chronic diseases found only limited improvements in general health status and clinical outcomes [14].

From a behavioral science perspective, effective management of people with chronic conditions requires more than just the use of mHealth technologies. It also requires consistent support from health care professionals over medium and long terms. In particular, studies focused on obesity have used interaction with professionals as an indicator of engagement. This interaction has been shown to play a significant role in achieving long-term (ie, 1 year) weight loss outcomes [15] although these definitions may vary between studies, potentially affecting the interpretation of results. Therefore, to maximize the potential of mHealth, we hypothesized that combining mHealth with health professional–led interventions would be more effective than relying solely on health care use. To investigate this hypothesis, our study focused exclusively on interventions that integrated mHealth with professional-led chronic disease strategies and then evaluated their effectiveness for each different duration.

Objectives

With this systematic review and meta-analysis, we investigated the effectiveness of combining mHealth and health professional–led intervention for improving health-related outcomes in chronic diseases.

Methods

Study Design

This was a systematic review and meta-analysis of randomized controlled trials.

Protocol and Registration

This systematic review was preregistered in the PROSPERO database (CRD42022337882) and conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1).

Search Strategy

Our search spanned 4 databases from their start dates until July 17, 2023. These included Web of Science, CENTRAL, MEDLINE, and CINAHL, as they are the four most used databases in similar research. Manual searches systematically screened the reference lists of all studies identified in the database search. The same inclusion criteria were applied to these references and the titles and abstracts were reviewed to determine whether they were included studies. Our systematic search strategy used both medical subject heading terms and keywords derived from subject headings. We limited our inclusion to randomized controlled trials published in English, excluding gray literature. The detailed search strategy is available in Multimedia Appendix 2.

Identification and Selection of Trials

Population

The population of interest included patients aged 18 years and older with at least 1 chronic condition. This was a change from the original protocol (CRD42022337882), which specified a minimum age of 20 years, in order to increase the generalizability of the findings by including all adults.

Interventions

This study's interventions comprised a mix of mHealth and direct health care professional involvement. The World Health Organization defines mHealth as "the use of mobile and wireless technologies to support health objectives" [16]. We excluded interventions that primarily used computers, categorizing them

as eHealth. Additionally, interventions where participants used smartphones solely for calls or text messages, without using specific apps, were not included to avoid overlapping mHealth and eHealth interventions and to ensure a clear distinction. The health care professionals involved in our study spanned various fields, including doctors, nurses, physiotherapists, occupational therapists, and dieticians. These professionals were actively engaged in delivering the mHealth intervention, beyond just explaining or reviewing app data. The methods used by health professionals for interventions included face-to-face interactions, phone calls, and text messaging.

Comparison

The comparison groups in this study consisted of the general control groups used in each individual study. For example, control groups were defined as those with only general care and follow-up, or those using mHealth-related devices but no involvement of a health care professional.

Outcomes

This systematic review assessed the following primary outcomes: HbA_{1c} levels, quality of life (QoL), and physical activity. These outcomes were chosen for their relevance in chronic disease management and were analyzed across different follow-up periods (short, medium, and long terms) to evaluate intervention effectiveness over time.

Selection Process

Rayyan software was used. Two authors (MK and TS) independently assessed the title and abstract records to identify eligible studies. The same authors then conducted a full-text screening to assess whether the criteria were met; in cases where the 2 authors' judgments conflicted, a third author (TM) decided and ultimately reached a consensus.

Assessment of Methodological Quality

The risk of bias for individual studies was assessed using the Cochrane Risk of Bias Framework for sequence generation, allocation concealment, blinding, completeness of outcome assessment, selective reporting, and other biases. Data were extracted by 2 authors (TM and MK). Disagreements were reviewed by a third author and a final consensus was reached. For the quality of the clinical evidence, the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) system was used to evaluate the results.

Data Extraction

Data were extracted by 2 authors (TM and MK). Disagreements were reviewed by a third reviewer and a final consensus was reached. Data were extracted on (1) basic information about the study, such as authors, year of publication, country, and region where the intervention took place; (2) target population; (3) intervention group description; (4) control group description; (5) outcomes; and (6) key findings.

Data Analysis

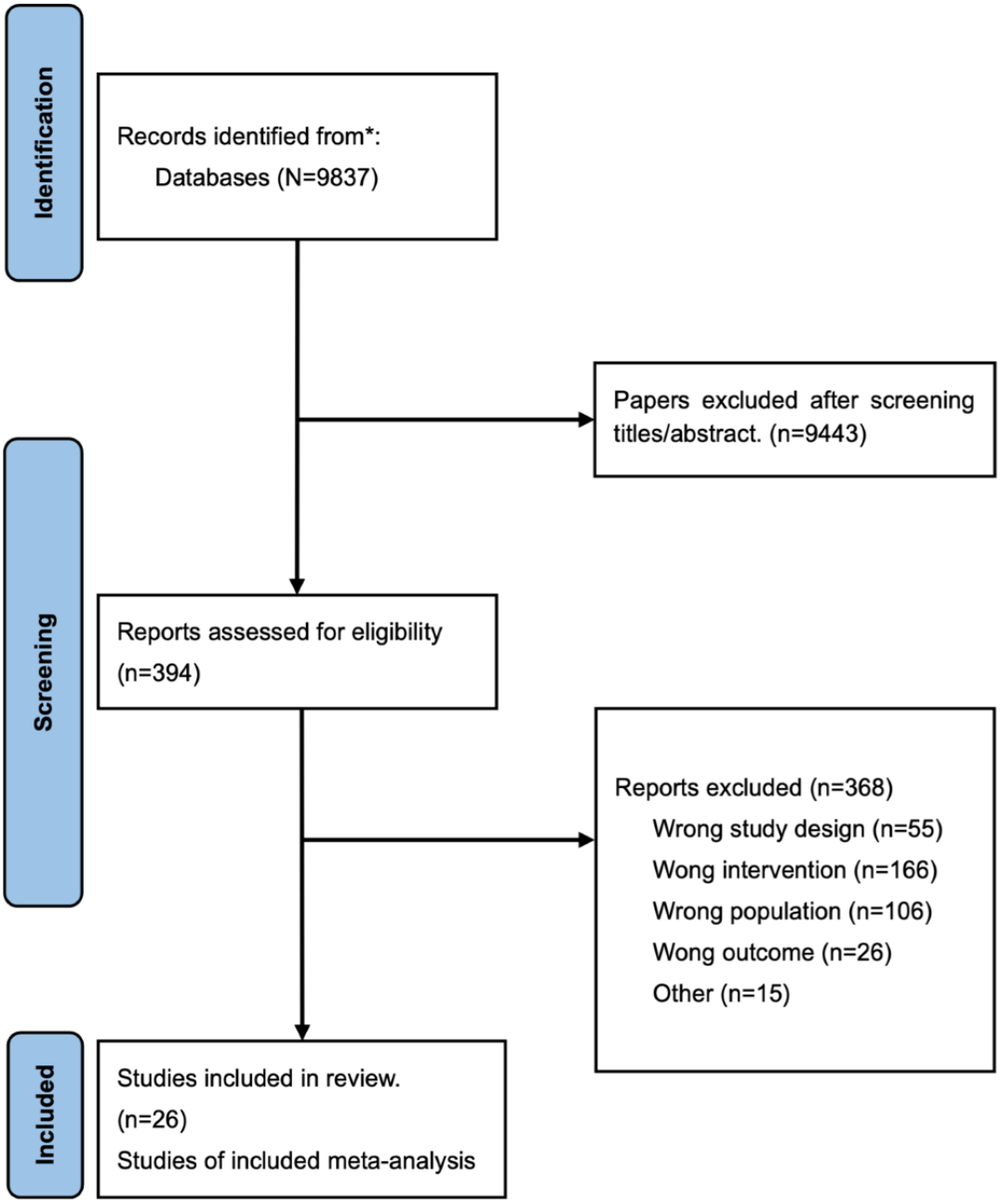
When two or more studies assessed the same outcome, a meta-analysis was performed using Review Manager 5.4 (Nordic Cochrane Centre). The results of the intervention groups were compared with the control group in each meta-analysis. The combined app and professional intervention was considered the intervention group and compared to the control group when there were more than 2 groups. The meta-analysis evaluated 3 different categories of time periods: short-, medium-, and long-term effects. Short term was defined as the closest to 3 months, medium term as the closest to 6 months, and long term as the closest to 12 months. For each follow-up point, the mean and SD were calculated. The group mean was calculated from the SE, CI, and *P* value of the difference between the means in trials that did not report SDs for each intervention group. In studies that only reported the median and IQR without providing SDs, we estimated the mean using the median value directly and approximated the SD by multiplying the IQR by 1.35. This method follows established meta-analytic conventions for handling summary statistics in studies with limited data [17]. All meta-analyses were presented as standardized mean difference (SMD) to account for adjusted and unadjusted means, due to the different data presentation formats of the included studies [17]. The authors of the studies in which the data were presented in a graphical format were contacted to obtain the exact values. If the reporting of results was still unclear after the above procedure, studies were excluded from the meta-analysis. For each meta-analysis, heterogeneity was assessed using I^2. When the I^2 value exceeded 75%, sensitivity analysis was conducted by sequentially excluding studies with the largest effect sizes, as these studies contributed significantly to the overall heterogeneity. This stepwise exclusion continued until heterogeneity dropped below the 75% threshold. By reducing the influence of studies with disproportionately high effect sizes, we aimed to create a more consistent and interpretable dataset, enhancing the reliability of the pooled effect size and minimizing the risk of bias or misinterpretation due to extreme variability among studies [18].

Results

Study Selection

The initial search identified 9837 records, which were screened for title and text as primary screening after removing duplicates. A number of 394 were selected for secondary screening after the first screening. Finally, 26 studies were finally selected. Figure 1 shows the flow chart of the selection process.

Figure 1. Flow diagram.



Characteristics of Included Trials and Intervention Description

Regarding the country, studies were conducted at outpatient or community practices in the following countries: Pakistan (2/26, 8%) [19,20], Spain (4/26, 15%) [21-24], Mexico (2/26, 8%) [25,26], United Kingdom (4/26, 15%) [27-30], United States (2/26, 8%) [31,32], China (4/26, 15%) [33-36], Singapore (2/26, 8%) [34,37]; and one each in Netherlands [38], Australia [39], Canada [40], Germany [41], Indonesia [42], and Thailand [43].

Regarding the disease, diabetes was the most common disease, with 12 of 26 trials [21,22,25,29-31,33,34,36,39,42,43]. Seven studies included type 2 diabetes [21,25,30,31,33,34,39] and 5 studies included both [22,29,36,42,43]. No study included only type 1 diabetes. The next most common was heart disease, which accounted for 7 of the 26 cases [19,20,24,35,37,38,44]. Other disease areas included cancer (2 studies) [26,32], respiratory diseases (1 study) [28], chronic diseases (2 studies) [23,40], metabolic syndrome (1 study) [41], and depression (1 study) [27].

The most common outcome was quality of life. There were 15 studies [20,22,24-27,30,32,34,37-41,44]. Ten studies used physical activity as an outcome [19,21,26,28,30,32,35,39,40,44]. Details are provided in Multimedia Appendix 3 [19-44].

Risk of Bias Assessment

The results of the risk of bias assessment are summarized in Multimedia Appendix 4 [19-44]. The risk of bias assessment for studies on HbA_{1c}, QoL, and physical activity (steps and subjective measures) showed that most studies had a low risk of bias. For example, studies like Anzaldo-Campos et al [25] and Blair et al [26] were evaluated as low-risk. Some studies had some concerns, such as Azelton et al [31] and Gill et al [40], but none were classified as high risk. Several studies, including

Franc et al [22] and Manzoor et al [19], did not have sufficient data for a full assessment.

Certainty of the Evidence According to the GRADE Approach

The certainty of evidence results using the GRADE approach are described in the results for each outcome. The GRADE assessment showed varying levels of evidence quality across outcomes. Most HbA_{1c} outcomes had very low to moderate certainty, with short-term outcomes showing very low certainty and medium-term outcomes showing moderate certainty. QoL outcomes generally had low to moderate certainty, with long-term outcomes showing very low certainty. Physical activity outcomes varied, with step counts showing moderate certainty and subjective measures ranging from very low to moderate certainty. These findings highlight the need for caution due to risks of bias and inconsistencies. A table with the results is provided in Multimedia Appendix 5.

Synthesis of Results: Meta-Analysis

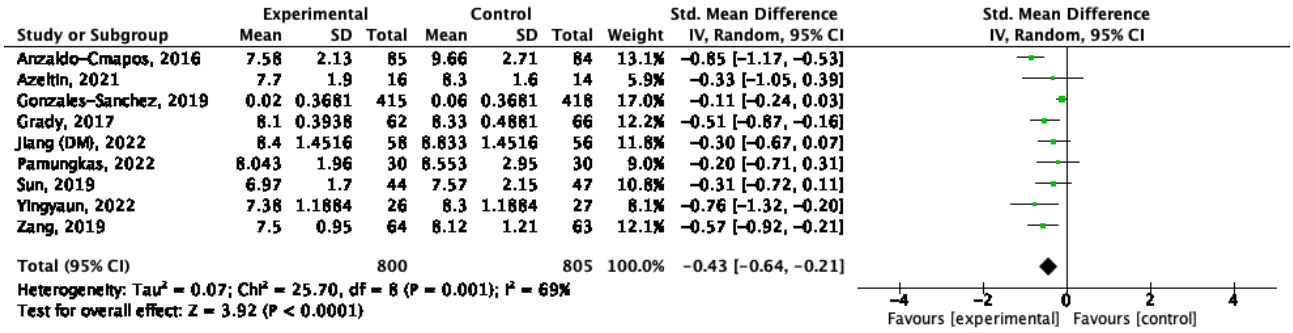
One study was excluded because data were not available [24]. Therefore, a meta-analysis was performed from 25 studies. Our systematic review included a broad spectrum of conditions, but diabetes was the most frequently addressed condition among the included studies. Consequently, HbA_{1c} was chosen as a primary indicator due to its importance in diabetes management and its frequent reporting in the studies. QoL was selected as it is a critical outcome measure reflecting the overall well-being of patients with chronic conditions. Physical activity and subjective outcomes of physical activity were included because these measures are commonly used to assess lifestyle modifications and their impact on chronic disease management. These indicators were the most frequently reported across the studies we reviewed, making them the most relevant for our analysis.

HbA_{1c} Outcome

Short Term

Meta-analysis was performed on 9 studies [23,25,29,31,33,34,36,42,43]. There was a significant increase in improvement for the combination intervention (SMD –0.43; 95% CI –0.64 to –0.21; P<.001; I^2=69%; Figure 2 [23,25,29,31,33,34,36,42,43]).

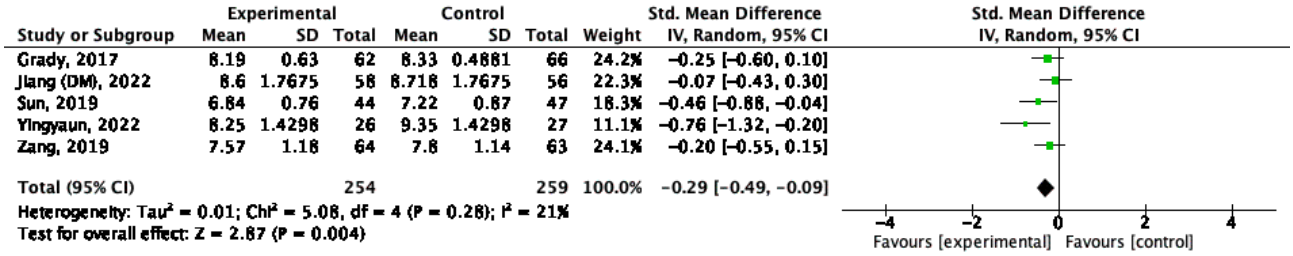
Figure 2. Hemoglobin A1c (HbA_{1c}) for short term [23,25,29,31,33,34,36,42,43].



Medium Term

A meta-analysis was performed on 5 studies [29,33,34,36,43]. There was a significant increase in improvement for the combination intervention (SMD –0.49; 95% CI –0.49 to –0.09; P<.001; I^2=21%; Figure 3 [29,33,34,36,43]).

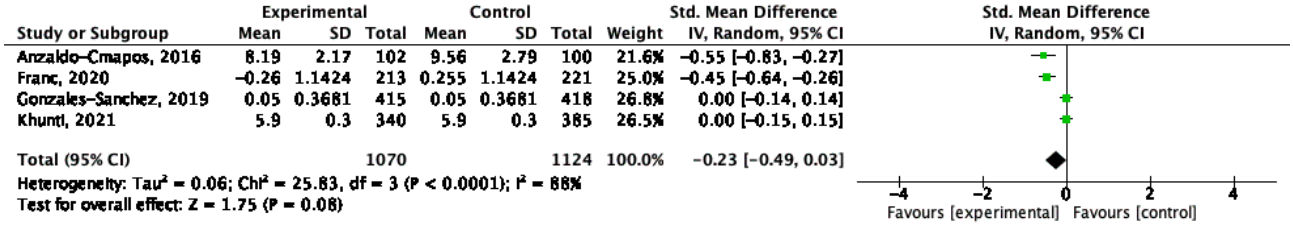
Figure 3. Hemoglobin A1c (HbA_{1c}) for medium term [29,33,34,36,43].



Long Term

One study was excluded because data were not available [24]. A meta-analysis was performed on 4 studies [22,23,25,30]. There was no significant increase in improvement for the combination intervention in the long term (SMD –0.23; 95% CI –0.49 to 0.03; P<.001; I^2=88%; Figure 4 [22,23,25,30]). Sensitivity analysis was performed because heterogeneity was >75%. Two studies [23,30] were excluded and did not reach statistical significance (SMD –0.00; 95% CI –0.10 to 0.10; P<.001; I^2=0%).

Figure 4. Hemoglobin A1c (HbA_{1c}) for long term [22,23,25,30].

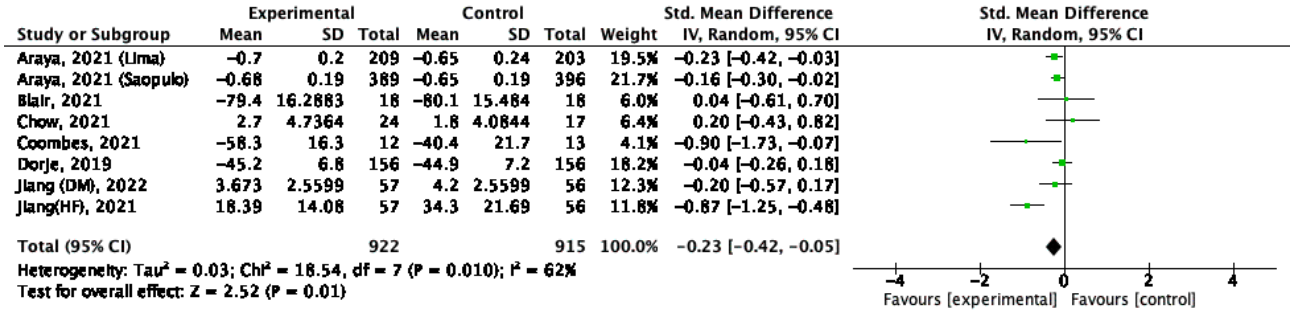


QoL Outcome

Short Term

Seven studies were included. However, 1 study [27] considered each of the 2 real samples, so they were included in the meta-analysis separately. Therefore, a meta-analysis was performed as 8 studies [26,27,32,34,37,39,44]. There was a significant increase in improvement for the combination intervention in the short term (SMD –0.23; 95% CI –0.42 to –0.05; P<.001; I^2=62%; Figure 5 [26,27,32,34,37,39,44]).

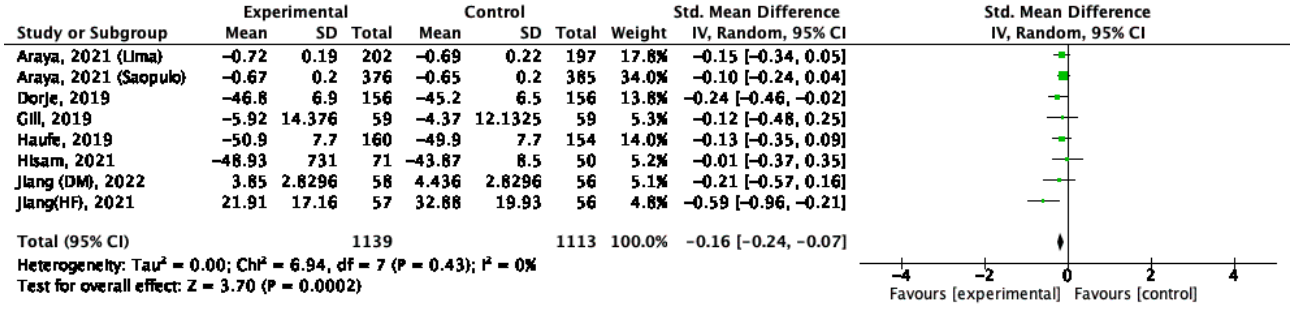
Figure 5. Quality of life for short term [26,27,32,34,37,39,44].



Medium Term

Seven studies were included. However, 1 study [27] considered each of the 2 real samples, so they were included in the meta-analysis separately. Therefore, a meta-analysis was performed as 8 studies [20,27,34,37,40,41,44]. There was a significant increase in improvement for the combination intervention (SMD –0.16; 95% CI –0.24 to –0.07; P<.001; I^2=0%; Figure 6 [20,27,34,37,40,41,44]).

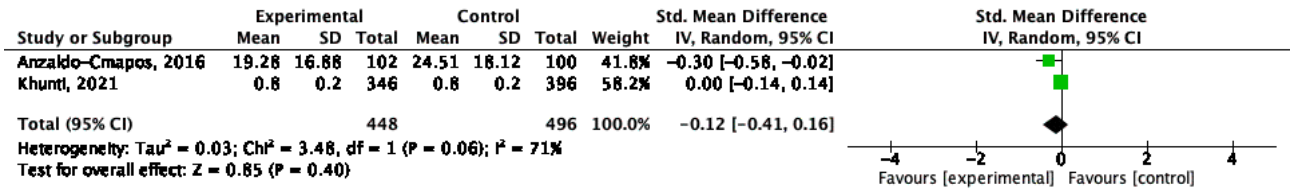
Figure 6. Quality of life for medium term [20,27,34,37,40,41,44].



Long Term

One study was excluded because data was not available [24]. A meta-analysis was performed on 2 studies [25,30]. There was no significant increase in improvement for the combination intervention (SMD –0.12; 95% CI –0.41 to 0.16; P=.04; I^2=71%; Figure 7 [25,30]).

Figure 7. Quality of life for long term [25,30].



Physical Activity

We performed a meta-analysis of the outcome of physical activity, dividing the data into subjective (questionnaire) and objective (number of steps).

Objective Outcomes of Physical Activity

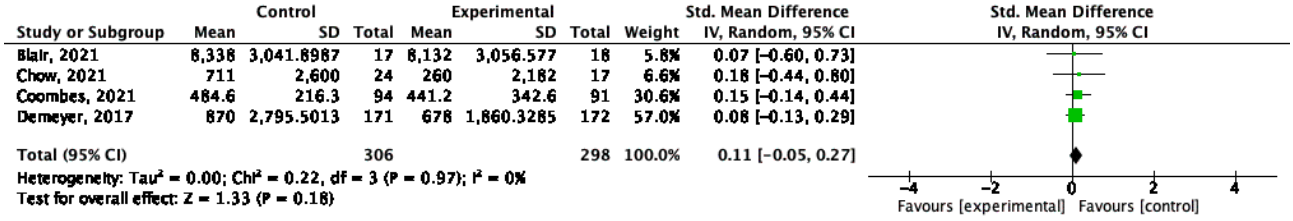
Overview

It was not possible to perform a meta-analysis for the medium and long terms because there was less than 1 study that corresponded to this time.

Short Term

Meta-analysis was conducted in 4 studies [26,28,32,39]. There was no significant increase in improvement for the combination intervention (SMD 0.11; 95% CI –0.05 to 0.27; P=.18; I^2=0%; Figure 8 [26,28,32,39]).

Figure 8. Physical activity (steps) for short term [26,28,32,39].

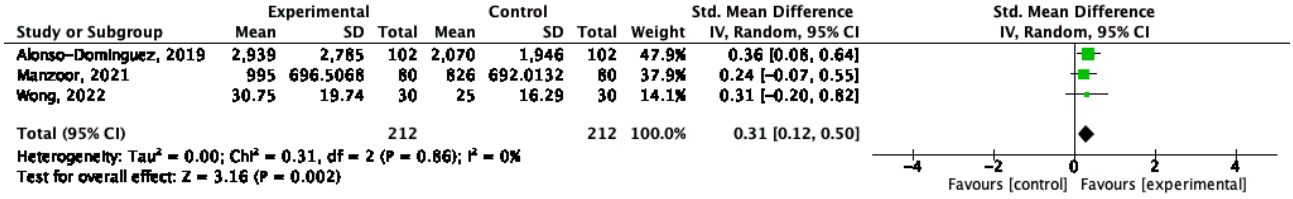


Subjective Outcomes of Physical Activity

Short Term

A meta-analysis of 3 studies was conducted [19,21,35]. There was a significant increase in improvement for the combination intervention (SMD 0.31; 95% CI 0.12-0.50; P=.002; I^2=0%; Figure 9 [19,21,35]).

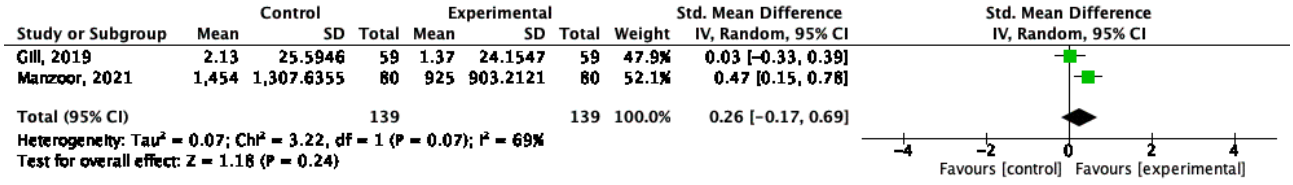
Figure 9. Physical activity (subjective) for short term [19,21,35].



Medium Term

Meta-analysis was conducted in 2 studies [19,40]. There was no significant increase in improvement for the combination intervention (SMD 0.26; 95% CI –0.17 to 0.69; P=.24; I^2=69%; Figure 10 [19,40]).

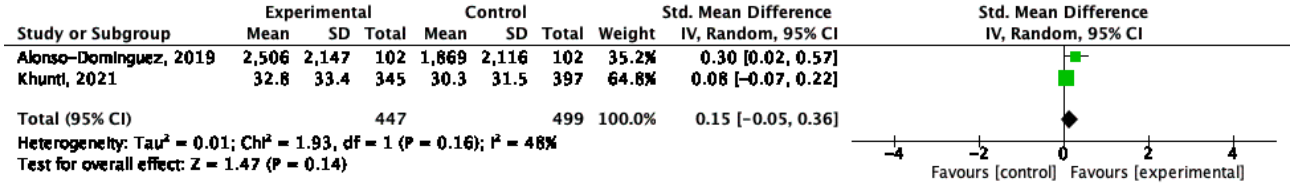
Figure 10. Physical activity (subjective) for medium term [19,40].



Long Term

Meta-analysis was conducted in 2 studies [21,30]. There was no significant increase in improvement for the combination intervention (SMD 0.15; 95% CI –0.05 to 0.36; P=.14; I^2=48%; Figure 11 [21,30]).

Figure 11. Physical activity (subjective) for long term [21,30].



Discussion

Principal Findings

This systematic review and meta-analysis showed that the combination of mHealth and health professional–led interventions improved health-related outcomes such as HbA_{1c} and QoL in chronic diseases. These effects were significant in the short and medium terms and there was no significant long term. For physical activity, there were no differences in intervention effects except for the short term as measured by questionnaires.

HbA_{1c}

In this systematic review, studies targeting patients with chronic diseases were included, and among the 26 studies, 12 incorporated HbA_{1c} as an outcome measure. In a meta-analysis that used mobile apps for lifestyle modification among diabetes patients, similar to our study, improvements in HbA_{1c} levels were demonstrated [45]. Furthermore, the meta-analysis indicated significant long-term effects in its subgroup analysis, defining long-term as 9-12 months [45], which differs from our study's definition. The impact of mHealth was indicated to diminish in the long term due to factors such as adherence fatigue [46]. However, another meta-analysis found that digital health–led diabetes self-management education and support (including mHealth, eHealth, and interventions using social networking services) were effective in improving HbA_{1c} at 6 and 12 months [47]. The discrepancies between our study and the other meta-analysis may be due to some factors. These factors include variations in digital health interventions, professional involvement levels, patient populations, and adherence. The other meta-analysis also included mHealth, eHealth, and interventions using social networking services, which may offer different engagement and support levels. These differences in intervention design and implementation could contribute to the varying long-term effectiveness observed. An overview of the studies we included in our systematic review, for example, Anzaldo-Campos et al [25] demonstrated that a smartphone app used for diabetes management significantly improved HbA_{1c} levels over a 6-month period, highlighting the potential of app-based interventions for glycemic control. In contrast, Franc et al [22] found that a telemonitoring system improved HbA_{1c} levels significantly over 12 months, suggesting that the integration of professional support enhances the effectiveness of digital health interventions. The differing results could be due to the varying levels of professional involvement and continuous monitoring provided in the telemonitoring system, which may help sustain patient adherence and engagement over a longer period. Additionally, Jiang et al [37] found that a nurse-led smartphone intervention significantly reduced HbA_{1c} levels over 6 months in patients with type 2 diabetes, emphasizing the importance of professional guidance in mHealth interventions. Conversely, Zhang et al [36] reported that a self-guided mobile app intervention without professional support did not result in significant improvements in HbA_{1c} levels, highlighting the potential limitations of mHealth interventions lacking continuous professional involvement. Our findings suggest that while the combination of mHealth and health professional–led interventions can be as diverse as those described above, continuous feedback and support to patients through other digital interventions to enhance engagement may be a key factor in achieving long-term benefits [48].

QoL

QoL was the most prevalent outcome measure among the studies included in this study. The results of our meta-analysis suggested that combining mHealth and health professional–led intervention can significantly improve QoL in chronic diseases in the short and medium terms. Studies conducting meta-analyses on the QoL in mHealth interventions for patients with chronic diseases are limited [49-51]. In a meta-analysis by Qin et al [50], mobile app–based interventions improved the QoL (SMD 0.39; 95% CI 0.27-0.51; $P<.001$) of patients with cancer. They emphasized that short-term interventions in particular (duration of 3 months or less), physician-patient interaction interventions, and cognitive behavioral therapy–based interventions may be most effective in improving QoL, which was consistent with this study. While extending beyond the scope of mHealth interventions, there is only 1 reported meta-analysis that indicated long-term improvements in QoL (mean difference 0.92; 95% CI 0.06-1.78; $P=.04$) resulting from cardiac telerehabilitation for patients with coronary artery disease [52]. For instance, Araya et al [27] found that a digital intervention combined with professional support significantly improved QoL in patients with cardiovascular diseases over 12 months. This improvement could be attributed to the continuous monitoring and personalized feedback provided by health care professionals, which likely enhanced patient adherence and engagement with the intervention. In contrast, Blasco et al [24] reported improvements in QoL among patients with acute coronary syndrome using a web-based telemonitoring system. The significant improvements observed in Blasco et al [24] study might be due to the comprehensive nature of the telemonitoring system, which included regular remote check-ins and the ability to promptly address patient concerns, thereby providing a sense of security and continuous care. Similarly, Manzoor et al [19] demonstrated that the mHealth-augmented cardiac rehabilitation program significantly improved QoL in patients with postacute coronary syndrome over 24 weeks, underscoring the importance of combining digital interventions with structured professional support. On the other hand, Wong et al [35] found that a self-guided mHealth intervention without professional involvement had limited effects on improving QoL, indicating that professional support is crucial for maximizing the benefits of digital health interventions. In our meta-analysis, focusing solely on mHealth interventions, it might be possible to identify long-term effects on the improvement of QoL by expanding the scope to include telehealth intervention or telerehabilitation as well.

Physical Activity

The results of our meta-analysis suggested that physical activity had no significant effect except for the short-term effect of subjective assessment. One reason for this might be that many mHealth interventions did not include specifications or support

to promote physical activity. This could be attributed to the predominant focus of many interventions on monitoring blood data and providing advice [22,24,25,29,31,33,34,36,42,43], potentially falling short of implementing behavior changes that could lead to promoting physical activity. In fact, among the 26 studies, only 8 studies adopted physical activity as the primary outcome. As an example of a study that adopted physical activity as the primary outcome, Gill et al [40] examined the effects of using a mobile app focused primarily on increasing physical activity and professional support for participants at risk for chronic disease and reported not only an increase in steps over the next 6 months but also maintenance of steps after 1 year. Dorje et al [44] found that a mHealth intervention including exercise modules improved physical activity levels in patients with coronary heart disease over 12 months, highlighting the potential of structured digital interventions. Conversely, Demeyer et al [28] reported that a self-guided digital intervention without professional support did not significantly increase physical activity levels in patients with chronic obstructive pulmonary disease, suggesting the need for continuous professional involvement to achieve sustained behavioral changes. In a different disease from the above, Blair et al [26] showed that a combination of mHealth and professional-led interventions for patients with cancer increased the number of steps taken, although there was no effect on outcomes related to sedentary behavior among physical activity measures. Thus, it would be valuable to identify in future studies which chronic diseases are more likely to improve physical activity with a combination of mHealth and professional-led interventions. Stavric et al [53] showed that digital physical activity and exercise interventions had positive effects on self-reported physical activity in people with chronic diseases, but not on objectively measured physical activity, which supports the results of our study. Their meta-analysis included self-guided digital mHealth interventions [53]. Therefore, there is room for future research to investigate whether self-directed digital or mHealth interventions or a combination of mHealth and health professional–led interventions, have more positive effects on physical activity.

Implications for Practice

The results of our study suggest that combining mHealth with professionally led interventions can be effective in chronic disease management, particularly in short and medium terms. This combination approach shows potential for improving important health outcomes such as HbA_{1c} and QoL. However, the lack of long-term effects indicates the need for sustained and adaptive intervention strategies. Health care providers should consider integrating ongoing digital support and feedback mechanisms to maintain patient engagement and adherence over time. In addition, the involvement of health care professionals in mHealth interventions should be designed to provide personalized and digital support, leveraging technologies such as telemedicine and mobile apps. For example, health care organizations should invest in training and infrastructure to support the seamless integration of mHealth technologies into traditional care models. This can improve the scalability and effectiveness of chronic disease management programs, ultimately leading to better patient outcomes and reduced health care costs.

Limitations

First, a major limitation of our study is the ambiguity in the definition of "health professional–led intervention." Our research focused on a combination of such interventions with mHealth strategies. However, the term may be misleading, as many studies typically include some level of health professional involvement in routine care. In contrast, our study specifically included cases with a more active role for health professionals than is typically seen in routine care. Additionally, the definition of the control group varied across studies. Some studies included only general care, while others provided mHealth devices without additional interventions. Second, our focus on health-related outcomes in chronic diseases limited our ability to conduct comprehensive meta-analyses due to the diverse nature of these diseases and their outcomes. Finally, we did not assess the cost-effectiveness of integrating health professional–led interventions with mHealth strategies. While the involvement of health professionals could potentially improve intervention effectiveness and reduce dropout rates, it could also lead to higher costs. Although mHealth interventions are known to struggle with high dropout rates [54], the cost implications of incorporating health professional support need to be further explored in future studies.

Conclusions

This systematic review and meta-analysis showed that the combined mHealth and health professional–led intervention had positive effects on HbA_{1c} and QoL in chronic conditions in short and medium terms. Positive effects on physical activity were only observed in the short term as measured by questionnaires. To achieve long-term effects of more than 1 year, it may be necessary to implement digital interventions that provide continuous feedback and support, tailored to a combination of mHealth and health professional–led interventions. Future research should explore the sustainability of these interventions over extended periods and investigate the specific components of mHealth and health professional input that contribute most to positive outcomes. Additionally, the integration of advanced technologies such as artificial intelligence and machine learning could further enhance the personalization and effectiveness of mHealth interventions.

Acknowledgments

The authors are grateful to Kono and Nemoto of Hokkaido University Library for their assistance in the development of the search formula.

Conflicts of Interest

PREVENT Inc is a company providing mHealth-based disease management programs. MK has received consulting fees from PREVENT Inc and is a nonregular staff member. TM and TS are employees of PREVENT Inc. YH is a founder and stockholder of PREVENT Inc.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - ijmr_v14i1e55835_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 30 KB - ijmr_v14i1e55835_app2.docx]

Multimedia Appendix 3
Characteristics of included studies with author names beginning with A to Z.
[DOCX File , 32 KB - ijmr_v14i1e55835_app3.docx]

Multimedia Appendix 4
Risk of bias assessment.
[DOCX File , 66 KB - ijmr_v14i1e55835_app4.docx]

Multimedia Appendix 5
Certainty of the evidence according to the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach.
[DOCX File , 20 KB - ijmr_v14i1e55835_app5.docx]

References

1. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15:1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
2. WHO Global Observatory for eHealth. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. Geneva, Switzerland: World Health Organization; 2011.
3. Kassavou A, Wang M, Mirzaei V, Shpendi S, Hasan R. The association between smartphone app-based self-monitoring of hypertension-related behaviors and reductions in high blood pressure: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(7):e34767 [FREE Full text] [doi: 10.2196/34767] [Medline: 35819830]
4. Bollyky JB, Bravata D, Yang J, Williamson M, Schneider J. Remote lifestyle coaching plus a connected glucose meter with certified diabetes educator support improves glucose and weight loss for people with type 2 diabetes. J Diabetes Res 2018;2018:3961730 [FREE Full text] [doi: 10.1155/2018/3961730] [Medline: 29888288]
5. Kanai M, Toda T, Yamamoto K, Akimoto M, Hagiwara Y. A mobile health–based disease management program improves blood pressure in people with multiple lifestyle-related diseases at risk of developing vascular disease—a retrospective observational study. Circ Rep 2022;4(7):322-329 [FREE Full text] [doi: 10.1253/circrep.CR-22-0024] [Medline: 35860354]
6. Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. Prev Med Rep 2018;12:284-293 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
7. Schmidt H. Chronic disease prevention and health promotion. In: Barrett DH, Ortmann LW, Dawson A, Saenz C, Reis R, Bolan G, editors. Public Health Ethics: Cases Spanning the Globe. Cham, Switzerland: Springer; 2016.
8. Tucker S, Bramante C, Conroy M, Fitch A, Gilden A, Wittleder S, et al. The most undertreated chronic disease: addressing obesity in primary care settings. Curr Obes Rep 2021;10(3):396-408 [FREE Full text] [doi: 10.1007/s13679-021-00444-y] [Medline: 34297343]
9. Kario K, Nomura A, Harada N, Okura A, Nakagawa K, Tanigawa T, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021;42(40):4111-4122 [FREE Full text] [doi: 10.1093/eurheartj/ehab559] [Medline: 34455443]
10. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
11. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]

12. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
13. Yi JY, Kim Y, Cho YM, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
14. Castelyn G, Laranjo L, Schreier G, Gallego B. Predictive performance and impact of algorithms in remote monitoring of chronic conditions: a systematic review and meta-analysis. Int J Med Inform 2021;156:104620 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104620] [Medline: 34700194]
15. Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A. The relationship between weight loss outcomes and engagement in a mobile behavioral change intervention: retrospective analysis. JMIR Mhealth Uhealth 2021;9(11):e30622 [FREE Full text] [doi: 10.2196/30622] [Medline: 34747706]
16. mHealth: Use of appropriate digital technologies for public health. World Health Organization. 2018. URL: https://apps.who.int/gb/ebwha/pdf_files/wha71/a71_20-en.pdf [accessed 2024-12-17]
17. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: John Wiley & Sons; 2019.
18. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
19. Manzoor S, Hisam A, Aziz S, Mashhadi SF, Haq ZU. Effectiveness of mobile health augmented cardiac rehabilitation on behavioural outcomes among post-acute coronary syndrome patients: a randomised controlled trial. J Coll Physicians Surg Pak 2021;31(10):1148-1153. [doi: 10.29271/jcpsp.2021.10.1148] [Medline: 34601832]
20. Hisam A, Haq ZU, Aziz S, Doherty P, Pell J. Effectiveness of mobile health augmented cardiac rehabilitation (MCard) on health-related quality of life among post-acute coronary syndrome patients: a randomized controlled trial. Pak J Med Sci 2022;38:716-723 [FREE Full text] [doi: 10.12669/pjms.38.3.4724] [Medline: 35480536]
21. Alonso-Domínguez R, Patino-Alonso MC, Sánchez-Aguadero N, García-Ortiz L, Recio-Rodríguez JI, Gómez-Marcos MA. Effect of a multifactorial intervention on the increase in physical activity in subjects with type 2 diabetes mellitus: a randomized clinical trial (EMID study). Eur J Cardiovasc Nurs 2019;18(5):399-409. [doi: 10.1177/1474515119835048] [Medline: 30808196]
22. Franc S, Hanaire H, Benhamou PY, Schaepelynck P, Catargi B, Farret A, et al. DIABEO system combining a mobile app software with and without telemonitoring versus standard care: a randomized controlled trial in diabetes patients poorly controlled with a basal-bolus insulin regimen. Diabetes Technol Ther 2020;22(12):904-911 [FREE Full text] [doi: 10.1089/dia.2020.0021] [Medline: 32407148]
23. Gonzalez-Sanchez J, Recio-Rodriguez JI, Fernandez-delRio A, Sanchez-Perez A, Magdalena-Belio JF, Gomez-Marcos MA, et al. Using a smartphone app in changing cardiovascular risk factors: a randomized controlled trial (EVIDENT II study). Int J Med Inform 2019;125:13-21. [doi: 10.1016/j.ijmedinf.2019.02.004] [Medline: 30914176]
24. Blasco A, Carmona M, Fernández-Lozano I, Salvador C, Pascual M, Sagredo P, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
25. Anzaldo-Campos MC, Contreras S, Vargas-Ojeda A, Menchaca-Díaz R, Fortmann A, Philis-Tsimikas A. Dulce wireless Tijuana: a randomized control trial evaluating the impact of project Dulce and short-term mobile technology on glycemic control in a family medicine clinic in Northern Mexico. Diabetes Technol Ther 2016;18(4):240-251 [FREE Full text] [doi: 10.1089/dia.2015.0283] [Medline: 26914371]
26. Blair CK, Harding E, Wiggins C, Kang H, Schwartz M, Tarnower A, et al. A home-based mobile health intervention to replace sedentary time with light physical activity in older cancer survivors: randomized controlled pilot trial. JMIR Cancer 2021;7(2):e18819 [FREE Full text] [doi: 10.2196/18819] [Medline: 33847588]
27. Araya R, Menezes PR, Claro HG, Brandt LR, Daley KL, Quayle J, et al. Effect of a digital intervention on depressive symptoms in patients with comorbid hypertension or diabetes in Brazil and Peru: two randomized clinical trials. JAMA 2021;325(18):1852-1862 [FREE Full text] [doi: 10.1001/jama.2021.4348] [Medline: 33974019]
28. Demeyer H, Louvaris Z, Frei A, Rabinovich RA, de Jong C, Gimeno-Santos E, et al. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 2017;72(5):415-423 [FREE Full text] [doi: 10.1136/thoraxjnl-2016-209026] [Medline: 28137918]
29. Grady M, Katz LB, Cameron H, Levy BL. Diabetes app-related text messages from health care professionals in conjunction with a new wireless glucose meter with a color range indicator improves glycemic control in patients with type 1 and type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017;2(2):e19 [FREE Full text] [doi: 10.2196/diabetes.7454] [Medline: 30291092]
30. Khunti K, Griffin S, Brennan A, Dallosso H, Davies M, Eborall H, et al. Behavioural interventions to promote physical activity in a multiethnic population at high risk of diabetes: PROPELS three-arm RCT. Health Technol Assess 2021;25(77):1-190 [FREE Full text] [doi: 10.3310/hta25770] [Medline: 34995176]

31. Azelton KR, Crowley AP, Vence N, Underwood K, Morris G, Kelly J, et al. Digital health coaching for type 2 diabetes: randomized controlled trial of healthy at home. Front Digit Health 2021;3:764735 [FREE Full text] [doi: 10.3389/fdgth.2021.764735] [Medline: 34901926]
32. Chow EJ, Doody DR, Di C, Armenian SH, Baker KS, Bricker JB, et al. Feasibility of a behavioral intervention using mobile health applications to reduce cardiovascular risk factors in cancer survivors: a pilot randomized controlled trial. J Cancer Surviv 2021;15(4):554-563 [FREE Full text] [doi: 10.1007/s11764-020-00949-w] [Medline: 33037989]
33. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone–based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
34. Jiang Y, Ramachandran HJ, Teo JYC, Leong FL, Lim ST, Nguyen HD, et al. Effectiveness of a nurse-led smartphone-based self-management programme for people with poorly controlled type 2 diabetes: a randomized controlled trial. J Adv Nurs 2022;78(4):1154-1165. [doi: 10.1111/jan.15178] [Medline: 35170786]
35. Wong EML, Leung DYP, Tam HL, Ko SY, Leung AYM, Lam SC, et al. Effectiveness of a nurse-led support programme using a mobile application versus phone advice on patients at risk of coronary heart disease - a pilot randomized controlled trial. Risk Manag Healthc Policy 2022;15:597-610 [FREE Full text] [doi: 10.2147/RMHP.S355554] [Medline: 35422666]
36. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app–based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019;21(12):e15401 [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
37. Jiang Y, Koh KWL, Ramachandran HJ, Nguyen HD, Lim DS, Tay YK, et al. The effectiveness of a nurse-led home-based heart failure self-management programme (the HOM-HEMP) for patients with chronic heart failure: a three-arm stratified randomized controlled trial. Int J Nurs Stud 2021;122:104026. [doi: 10.1016/j.ijnurstu.2021.104026] [Medline: 34271265]
38. Broers ER, Widdershoven J, Denollet J, Lodder P, Kop WJ, Wetzels M, et al. Personalized eHealth Program for life-style change: results from the "Do Cardiac Health Advanced New Generated Ecosystem (Do CHANGE 2)" randomized controlled trial. Psychosom Med 2020;82(4):409-419. [doi: 10.1097/PSY.0000000000000802] [Medline: 32176191]
39. Coombes JS, Keating SE, Mielke GI, Fassett RG, Coombes BK, O'Leary KP, et al. Personal activity intelligence e-Health program in people with type 2 diabetes: a pilot randomized controlled trial. Med Sci Sports Exerc 2022;54(1):18-27. [doi: 10.1249/MSS.0000000000002768] [Medline: 34334715]
40. Gill DP, Blunt W, Boa Sorte Silva NC, Stiller-Moldovan C, Zou GY, Petrella RJ. The HealtheSteps™ lifestyle prescription program to improve physical activity and modifiable risk factors for chronic disease: a pragmatic randomized controlled trial. BMC Public Health 2019;19(1):841 [FREE Full text] [doi: 10.1186/s12889-019-7141-2] [Medline: 31253112]
41. Haufe S, Kerling A, Protte G, Bayerle P, Stenner HT, Rolff S, et al. Telemonitoring-supported exercise training, metabolic syndrome severity, and work ability in company employees: a randomised controlled trial. Lancet Public Health 2019;4(7):e343-e352 [FREE Full text] [doi: 10.1016/S2468-2667(19)30075-1] [Medline: 31204284]
42. Pamungkas RA, Usman AM, Chamroonsawasdi K, Abdurrasyid. A smartphone application of diabetes coaching intervention to prevent the onset of complications and to improve diabetes self-management: a randomized control trial. Diabetes Metab Syndr 2022;16(7):102537. [doi: 10.1016/j.dsx.2022.102537] [Medline: 35724489]
43. Yingyaun K, Charoensri S, Pongchaiyakul C. Effect of using mobile messenger for insulin injection education for glycemic control: a randomized controlled trial. Pac Rim Int J Nurs Res 2022;26:461-473 [FREE Full text]
44. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
45. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
46. Lee JJN, Abdul Aziz A, Chan S, Raja Abdul Sahrizan RSFB, Ooi AYY, Teh YT, et al. Effects of mobile health interventions on health-related outcomes in older adults with type 2 diabetes: a systematic review and meta-analysis. J Diabetes 2023;15(1):47-57 [FREE Full text] [doi: 10.1111/1753-0407.13346] [Medline: 36649940]
47. Nkhoma DE, Soko CJ, Bowrin P, Manga YB, Greenfield D, Househ M, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
48. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital Diabetes Prevention Program: 3-year update. BMJ Open Diabetes Res Care 2017;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
49. Chin-Jung L, Hsiao-Yean C, Yeu-Hui C, Kuan-Chia L, Hui-Chuan H. Effects of mobile health interventions on improving glycemic stability and quality of life in patients with type 1 diabetes: a meta-analysis. Res Nurs Health 2021;44(1):187-200. [doi: 10.1002/nur.22094] [Medline: 33368403]
50. Qin M, Chen B, Sun S, Liu X. Effect of mobile phone app–based interventions on quality of life and psychological symptoms among adult cancer survivors: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(12):e39799 [FREE Full text] [doi: 10.2196/39799] [Medline: 36534460]

51. Gong Y, Jiang X, Chen X, Chen S, Wen Y, Yuan X, et al. Effectiveness of mHealth diet interventions in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Asia Pac J Oncol Nurs 2023;10(3):100196 [FREE Full text] [doi: 10.1016/j.apjon.2023.100196] [Medline: 37124242]
52. Zhong W, Liu R, Cheng H, Xu L, Wang L, He C, et al. Longer-term effects of cardiac telerehabilitation on patients with coronary artery disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023;11:e46359 [FREE Full text] [doi: 10.2196/46359] [Medline: 37505803]
53. Stavric V, Kayes NM, Rashid U, Saywell NL. The effectiveness of self-guided digital interventions to improve physical activity and exercise outcomes for people with chronic conditions: a systematic review and meta-analysis. Front Rehabil Sci 2022;3:925620 [FREE Full text] [doi: 10.3389/fresc.2022.925620] [Medline: 36188933]
54. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]

Abbreviations

GRADE: Grading of Recommendations Assessment, Development, and Evaluation
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
QoL: quality of life
SMD: standardized mean difference

Edited by T de Azevedo Cardoso; submitted 27.12.23; peer-reviewed by L Benning, M Fraser; comments to author 23.05.24; revised version received 19.06.24; accepted 03.12.24; published 20.01.25.

Please cite as:
Kanai M, Miki T, Sakoda T, Hagiwara Y
The Effect of Combining mHealth and Health Professional–Led Intervention for Improving Health-Related Outcomes in Chronic Diseases: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e55835
URL: https://www.i-jmr.org/2025/1/e55835
doi:10.2196/55835
PMID:

Review

Comparing Digital Versus Face-to-Face Delivery of Systemic Psychotherapy Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Pieter Erasmus[1*], MA; Moritz Borrmann[1,2*], MA; Jule Becker[1], BA; Lars Kuchinke[3], PhD; Gunther Meinlschmidt[1,4,5,6], PhD

[1]Department of Clinical Psychology and Cognitive Behavioral Therapy, International Psychoanalytic University Berlin, Berlin, Germany

[2]Day Clinic for Cognitive Neurology, University Hospital Leipzig, Leipzig, Germany

[3]Psychological Methods and Evaluation, International Psychoanalytic University Berlin, Berlin, Germany

[4]Department of Clinical Psychology and Psychotherapy – Methods and Approaches, Trier University, Trier, Germany

[5]Department of Digital and Blended Psychosomatics and Psychotherapy, Psychosomatic Medicine, University Hospital and University of Basel, Basel, Switzerland

[6]Department of Psychology, University of Basel, Basel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Gunther Meinlschmidt, PhD
Department of Clinical Psychology and Cognitive Behavioral Therapy
International Psychoanalytic University Berlin
Stromstrasse 3b
Berlin, 10555
Germany
Phone: 49 30 300 117 710
Email: gunther.meinlschmidt@ipu-berlin.de

Abstract

Background: As digital mental health delivery becomes increasingly prominent, a solid evidence base regarding its efficacy is needed.

Objective: This study aims to synthesize evidence on the comparative efficacy of systemic psychotherapy interventions provided via digital versus face-to-face delivery modalities.

Methods: We followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for searching PubMed, Embase, Cochrane CENTRAL, CINAHL, PsycINFO, and PSYNDEX and conducting a systematic review and meta-analysis. We included randomized controlled trials comparing mental, behavioral, and somatic outcomes of systemic psychotherapy interventions using self- and therapist-guided digital versus face-to-face delivery modalities. The risk of bias was assessed with the revised Cochrane Risk of Bias tool for randomized trials. Where appropriate, we calculated standardized mean differences and risk ratios. We calculated separate mean differences for nonaggregated analysis.

Results: We screened 3633 references and included 12 articles reporting on 4 trials (N=754). Participants were youths with poor diabetic control, traumatic brain injuries, increased risk behavior likelihood, and parents of youths with anorexia nervosa. A total of 56 outcomes were identified. Two trials provided digital intervention delivery via videoconferencing: one via an interactive graphic interface and one via a web-based program. In total, 23% (14/60) of risk of bias judgments were *high risk*, 42% (25/60) were *some concerns*, and 35% (21/60) were *low risk*. Due to heterogeneity in the data, meta-analysis was deemed inappropriate for 96% (54/56) of outcomes, which were interpreted qualitatively instead. Nonaggregated analyses of mean differences and CIs between delivery modalities yielded mixed results, with superiority of the digital delivery modality for 18% (10/56) of outcomes, superiority of the face-to-face delivery modality for 5% (3/56) of outcomes, equivalence between delivery modalities for 2% (1/56) of outcomes, and neither superiority of one modality nor equivalence between modalities for 75% (42/56) of outcomes. Consequently, for most outcome measures, no indication of superiority or equivalence regarding the relative efficacy of either delivery modality can be made at this stage. We further meta-analytically compared digital versus face-to-face delivery modalities for attrition (risk ratio 1.03, 95% CI 0.52-2.03; P=.93) and number of sessions attended (standardized mean difference

–0.11; 95% CI –1.13 to –0.91; *P*=.83), finding no significant differences between modalities, while CIs falling outside the range of the minimal important difference indicate that equivalence cannot be determined at this stage.

Conclusions: Evidence on digital and face-to-face modalities for systemic psychotherapy interventions is largely heterogeneous, limiting conclusions regarding the differential efficacy of digital and face-to-face delivery. Nonaggregated and meta-analytic analyses did not indicate the superiority of either delivery condition. More research is needed to conclude if digital and face-to-face delivery modalities are generally equivalent or if—and in which contexts—one modality is superior to another.

Trial Registration: PROSPERO CRD42022335013; https://tinyurl.com/nprder8h

(Interact J Med Res 2025;14:e46441) doi:10.2196/46441

KEYWORDS

systemic psychotherapy; family therapy; adolescent; systematic review; meta-analysis; face to face; digital; remote; distance; telehealth; delivery modality

Introduction

Background

Digital delivery of mental health interventions has gained increasing prominence in recent decades. Digital delivery provides solutions to many mental health intervention obstacles, such as limited access [1], geographical distance [2], financial constraints [3], and transportation issues [4]. These factors, and not least the COVID-19 pandemic [5], have contributed to increased delivery of digital modalities of mental health care [6]. One current line of research uses randomized controlled trials (RCTs) to determine the efficacy of digitally delivered family therapies [6] and parenting interventions [7], including but not limited to systemic approaches. This research supports the efficacy of the digital delivery of interventions and finds no evidence for digitally delivered interventions being inferior to face-to-face interventions [8-13]. However, most of the current literature does not compare face-to-face and digitally delivered interventions directly, does not differentiate between the types of digital delivery modalities (eg, therapist-guided videoconferencing and self-guided web-based programs), and does not consider how the modality differentially impacts specific disorders, outcomes, or populations. Furthermore, these studies aggregate findings across broad and diverse contextual definitions of family therapy and parenting interventions instead of applying unified and stringent theory-informed definitions. Although these reviews (eg, McLean et al [6] and Florean et al [7]) do not exclusively focus on a unified and stringent definition, they do provide evidence for systemic approaches in the broad context of family interventions. Many family therapists consider themselves to be systemic therapists, and there is an overlap between family therapy and systemic therapy (ST). However, ST is grounded in a theoretical treatment model, and unlike family therapy, it is not primarily defined by the therapy setting [14]. Thus, ST provides a unified and stringent definition that has practical consequences for research and clinical practice.

ST is a conceptual framework for mental health interventions that focuses on interpersonal relations, interactions, social surroundings and resources, perspectives, and constructions of problems; attempted solutions are appreciated and used as integral parts of the intervention [13,15,16]. Consequently, most family therapy and parenting interventions can be said to contain elements of ST to varying degrees while also potentially including elements from other conceptual frameworks, such as cognitive behavioral therapy and psychoeducation [17]. Conversely, as ST interventions often focus on families as important resources requiring incorporation into intervention delivery [18], they can, but need not be, similar to other family therapy or parenting interventions in terms of including other family members in their setting. The efficacy of ST as a face-to-face intervention is well documented in various reviews and meta-analyses for a range of psychological disorders for youth and adult populations [13,15-17,19]. In line with developments for other psychotherapeutic interventions, the implementation of digital delivery modalities of ST has rapidly increased in recent years [20-22]. The implementation of videoconferencing is particularly promising, and the experiences of practitioners implementing this delivery modality are overall positive [23]. Despite such indications, there are gaps in the literature [20,24]. There is an urgent need to provide an evidence base for practitioners to understand how face-to-face and digital delivery of the same ST intervention compare (compare McLean et al [6], Florean et al [7], and Fairburn and Patel [25]).

The need for greater granularity regarding the comparative efficacy of digital and face-to-face delivery modalities is in line with the more general requirement to identify “conditions under which systemic therapy works best” [16]. Thus, we simultaneously address both the need for a systematic review and meta-analysis that investigates (1) how various types of ST interventions, delivered in their traditional face-to-face form, compare to the same ST intervention delivered digitally with regard to their efficacy and (2) delivery modality as key contextual factors that modulate ST intervention outcomes.

Research Questions

We conducted a systematic review and meta-analysis to address the following research questions: (1) What are the characteristics, quality, and resulting evidence from published RCTs comparing the efficacy of ST interventions using digital treatment modalities with ST interventions using face-to-face treatment modalities across settings, populations, and outcomes? (2) If a meta-analysis of RCTs comparing ST interventions in digital and face-to-face modalities can be conducted, what is the difference in delivery modalities for mental disorder outcomes?

Methods

Overview

This systematic review followed reporting guidelines as laid out in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement (Multimedia Appendix 1 2020 checklist) [26]. We aimed to identify, evaluate, and synthesize findings from RCTs comparing digital and face-to-face delivery modalities for ST interventions across settings, target populations, and outcomes. This systematic review was preregistered (PROSPERO ID: CRD42022335013), with no deviation from the registered protocol.

Inclusion and Exclusion Criteria

We applied the following inclusion and exclusion criteria for articles to be considered for our systematic review within the following categories: (1) study design, (2) participants, (3) interventions, (4) digital delivery and face-to-face delivery modalities, (5) outcomes, and (6) article type.

RCTs of any type were included (eg, parallel, cluster, and quasi-RCTs). No minimum number of participants was prespecified.

Only those articles were deemed eligible for inclusion that involved individual participants of all ages and genders (1) with one or more diagnoses of a mental disorder, (2) identified as at risk in one or more mental or behavioral health domains, or (3) adversely affected by salient circumstances (eg, belonging to a structurally disadvantaged community and caring for a family member). Articles involving participants at the system level (eg, families) were also included as long as 1 member of the system fulfilled at least one of the aforementioned criteria. No further exclusion criteria were applied to increase the number of eligible studies.

We only included articles reporting on interventions delivered via at least 2 distinct delivery modalities: digital delivery and face-to-face delivery. Digital delivery was defined as exclusively relying on an external electronic medium (eg, videoconferencing hardware and software, telephone, and synchronous and asynchronous text-based conversations) for intervention delivery. While this allowed heterogeneity of digital delivery modalities across some dimensions (eg, synchronicity and extent of interaction with mental health practitioners), it kept other dimensions stable across all digital delivery modalities (use of electronic medium and no mental health practitioner physically copresent with recipients of intervention). Face-to-face delivery was defined as exclusively relying on in-person intervention delivery. This allowed for the use of external tools or media to supplement the intervention while specifying that intervention delivery had to occur solely between people physically present in the same physical space (eg, a clinician's office).

To increase the number of eligible studies, any relevant mental, behavioral, or somatic health efficacy; adherence and attrition; and further outcomes were included in this review. No a priori categorization of outcomes was applied. Efficacy outcomes were defined as any outcomes pertinent to improvement in relevant mental, behavioral, and somatic health domains. *Further outcomes* were defined as any outcomes not directly pertinent to improvement in relevant health domains (eg, subjective satisfaction rating of program delivery).

There were no restrictions imposed on the article type in terms of language or date. Articles published in peer-reviewed journals as full-text articles and other types of peer-reviewed full-text articles were included. Unpublished data (eg, dissertations), as well as study protocols, clinical trial registries, and non–peer-reviewed articles, were excluded.

Intervention Inclusion Criteria

Only articles reporting on ST interventions delivered via both digital and face-to-face delivery modalities were included. ST was operationalized based on 4 adapted definition criteria of ST [13], which were applied to the descriptions of the identified constituent primary parts (CPPs; eg, core sessions) of interventions. The definition criteria were derived from the following definition of ST (referred to as *systemic therapy*) with our adjustments in square brackets [13]:

> *We use systemic/systems-oriented therapy/therapies (ST) as a general term for a major therapeutic orientation that can be distinguished from other major approaches (e.g., CBT or psychodynamic therapy). We define systemic therapy as a form of psychotherapy that (1) perceives behavior and mental symptoms within the context of the social systems people live in; (2) focuses on interpersonal relations and interactions, social constructions of realities, and[/or] the recursive causality between symptoms and interactions; (3) includes family members and[/or] other important persons (e.g., teachers, friends, professional helpers) directly or indirectly through systemic questioning, hypothesizing, and specific interventions; and (4) appreciates and utilizes clients' perspectives on problems, resources, and[/or] preferred solutions.*

For each intervention, the appropriate unit of primary intervention delivery (as opposed to parts of the intervention labeled *supplementary*, *additional*, or *optional*, etc) was determined following the study's authors (eg, core sessions, core components, etc.). The total number of the intervention's CPPs was determined. If the intervention description did not afford to determine an appropriate unit of intervention delivery, the total number of CPPs was defined as 1.

For each description of content for any CPP, we assessed if it met any of our definition criteria of ST. Notably, any particular definition criterion could, in theory, be met by a set of CPP descriptions >1, allowing different content descriptions to meet the same criterion. This allowed a range of different interventions to be classified as being similar in kind (ie, being ST interventions), reflecting the integrative nature of clinical ST practice [27,28], rather than requiring all interventions to have identical CPPs.

All definition criteria of ST needed to be met by the respective CPPs, with at least 2 of the 4 definition criteria being met completely by all CPPs. If any of the definition criteria of ST were not completely met, at most 2 of the 4 definition criteria needed to be met by a minimum of at least 50% of all CPPs (or,

in the case of uneven numbers of CPPs, the closest possible number below the theoretical half point). The same CPP could satisfy >1 criterion. Definition criterion (1) could additionally be met by the relevant conceptual background provided.

Textbox 1 illustrates an example for 1 trial included in our study. Our approach used to screen trials against inclusion criteria for all trials is detailed in Multimedia Appendix 2 [13,29-32].

Textbox 1. Identified constituent primary parts (CPPs) of 1 trial against systemic therapy (ST) definition criteria for Behavioral Family Systems Therapy for Diabetes (BFST-D) intervention [33].

Number and unit of CPPs and 4 primary intervention components

1. Perceives behavior and mental symptoms within the context of the social systems people live in
   - Family functioning and maladaptive parent-child interactions were identified as central barriers to diabetes treatment adherence.
2. Focuses on interpersonal relations and interactions, social constructions of realities, or the recursive causality between symptoms and interactions: 3 of 4 CPPs
   - CPP1: family-problem solving (as a family, defining the problem, generating solutions, making decisions, implementing and monitoring results, and refining ineffective solutions)
   - CPP2: communication training (instruction, feedback, modeling, and rehearsal of approaches toward improving maladaptive communicative patterns)
   - CPP4: family restructuring (functional and structural approaches toward changing maladaptive or ineffective family system patterns and characteristics, such as weak parental coalitions or cross-generational coalitions).
3. Includes family members and other important persons (eg, teachers, friends, professional helpers) directly or indirectly through systemic questioning, hypothesizing, and specific interventions
   - All CPPs are delivered to the caregiver-adolescent dyad.
4. Appreciates and uses clients' perspectives on problems, resources, and preferred solutions: 2 of 4 CPPs are as follows:
   - CPP1 and CPP3: cognitive restructuring (addressing beliefs, attitudes, and attributions that could negatively affect effective interactions).

Search Strategy

Database Search

The following databases were searched from the earliest available date until March 15, 2022: PubMed, Embase (via Ovid), Cochrane CENTRAL (via Ovid), CINAHL (via EBSCO), PsycINFO (via EBSCO), and PSYNDEX (via EBSCO). The search strategy followed the guideline from Bramer et al [34], and the original search string was developed for use in PubMed and adapted manually for all other databases. The search strings contained a set of the following: (1) 1 compiled term for intervention type, (2) 1 compiled specificity term to define the target domain, and (3) 1 compiled delivery modality term. For the identification of RCT study design, we used the RobotSearch AI RCT highly sensitive filter [35], which provides high sensitivity in identifying RCTs [35-37]. Multimedia Appendix 3 [8,13,15,38-46] provides the full search strings, references of the published sources upon which we based our search terms, and explanations of the logical structure. On March 17, 2022, we conducted complementary searches in ClinicalTrials.gov, the International Clinical Trials Registry Platform, and the EU Clinical Trials Register (all via the respective standard website interface). On May 4, 2022, backward citation screening (articles cited) was conducted on identified relevant reviews and meta-analyses (Multimedia Appendix 4 [9,10,12,13,15,47-53]). On May 10, 2022, forward citation screening (articles citing included articles) was conducted via ISI Web of Science, backward citation screening on included articles was conducted via PubMed, and a "related articles" screening was conducted on included articles (first 20 articles listed as related articles) via PubMed.

Article Selection and Screening

Duplicates were removed using Automated Systematic Search Deduplicator [54]. The resulting list of search results was split into 2 sets (sorting by title in EndNote [version 20; Clarivate] and allocating alternating sets of approximately 100 consecutive references; number of articles=1816 and 1817) that were each assigned to 2 pairs of authors (PE and JB; MB and JB), with each author independently screening titles and abstracts using manualized standard operating procedures that we developed accommodating our inclusion and exclusion criteria. All discrepancies at the title and abstract screening stage were resolved through discussion between all 3 involved authors (PE, MB, and JB). PE and MB independently screened at the full-text stage. Discrepancies at this stage were resolved through discussion between PE and MB. GM was consulted as an independent author of this study in cases of persisting discrepancies after discussions at any stage. Because on several occasions a trial underlying an included article contributed data to >1 included article, article-trial correspondence was determined either via trial registration numbers (for 7 articles) or via references to articles on the same trial, unique trial names, as well as overlap in article authors, article dates, sample sizes and characteristics, interventions, and procedures (for 5 articles).

Data Extraction

Data were independently extracted by PE and MB using the Cochrane data collection form [55] for intervention reviews

(Multimedia Appendix 5 [29,30,33,56-59]). Discrepancies were resolved through discussion.

Risk of Bias

The authors (PE and MB) independently assessed the risk of bias (RoB) using the revised Cochrane Risk of Bias (version 2) tool for randomized trials [60] and following the process outlined in chapter 8 of the Cochrane handbook [55]. Discrepancies were resolved through discussion. RoB figures were created using *robvis* [61].

Data Synthesis and Analysis

We used RevMan (version 5.4) [62] to create forest plots and calculate, depending on the variable type and type of analysis, mean differences (MDs), standardized MDs, risk ratios, and associated CIs at the 95% level. Where possible, we calculated missing SDs using the Cochrane collaboration RevMan calculator [62]. We conducted meta-analyses only in cases of acceptable levels of clinical and methodological diversity and where sufficient data were available [63]. We conducted meta-analyses using random effects models as we expected heterogeneity across articles to be high. Statistical heterogeneity was assessed using the I^2 statistic. MDs and CIs were calculated where meta-analytic analysis was not possible, but nonaggregated analysis was still deemed appropriate. This was to increase qualitative comparability across studies with nonnegligible heterogeneity in statistical approaches and reported results. All relevant reported results are listed in Multimedia Appendix 6 [29,30,33,56-59,64-68]. We grouped MDs of outcomes according to their CIs into (1) fully located within the range from –0.20 to 0.20 (*equivalence*), (2) fully located outside this range (*superiority*), or (3) partially located within and partially located outside this range (*neither equivalence nor superiority*). In the absence of an established a priori minimal important difference, we accept Cohen *d* cutoff of 0.2 (small effect) as the defined range from –0.20 to 0.20 to indicate equivalence [69]. We interpreted a CI of an MD falling within, but not exceeding, this range as evidence for equivalence, a CI of an MD falling entirely outside this range as evidence for nonequivalence or superiority of one delivery modality, and a CI falling within and exceeding this range as insufficient evidence to determine equivalence or nonequivalence on the respective measure. Where calculating MDs and CIs was impossible due to missing information, reported results on differences between digital and face-to-face delivery modalities were provided (Multimedia Appendix 6). Study authors were contacted to provide missing information. A funnel plot was not created because we identified <10 trials [70].

Results

Search Results

A flowchart of our search and screening process is depicted in Figure 1 [35]. The primary search for this study retrieved 4977 references and 3633 after removal of duplicates. Additional references (n=2) were identified through complimentary searches. The resulting 3635 references were screened at the title and abstract level. On the basis of our inclusion and exclusion criteria, 97.1% (3529/3635) articles were excluded during the title and abstract screening, and 2.9% (106/3635) articles were screened at the full-text screening stage. Overall, 0.3% (12/3633) of the initially identified articles reporting on 4 trials were included in this systematic review. Some (3/12, 25%) of the articles reporting on 2 trials [29,56,57] are only reported in Table 1 and Multimedia Appendix 6, as they could not be included in further meta-analytic aggregation or discrete analysis of MDs and CIs.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the articles included in the systematic review. RCT: randomized controlled trial. *Results from CINAHL (EBSCO): n=432; PSYNDEX (EBSCO): n=53; PsycINFO (EBSCO): n=505; Embase (Ovid): n=2177; Cochrane (Wiley): n=533; PubMed: n=4159; **RobotSearch AI, Marshall et al [35].

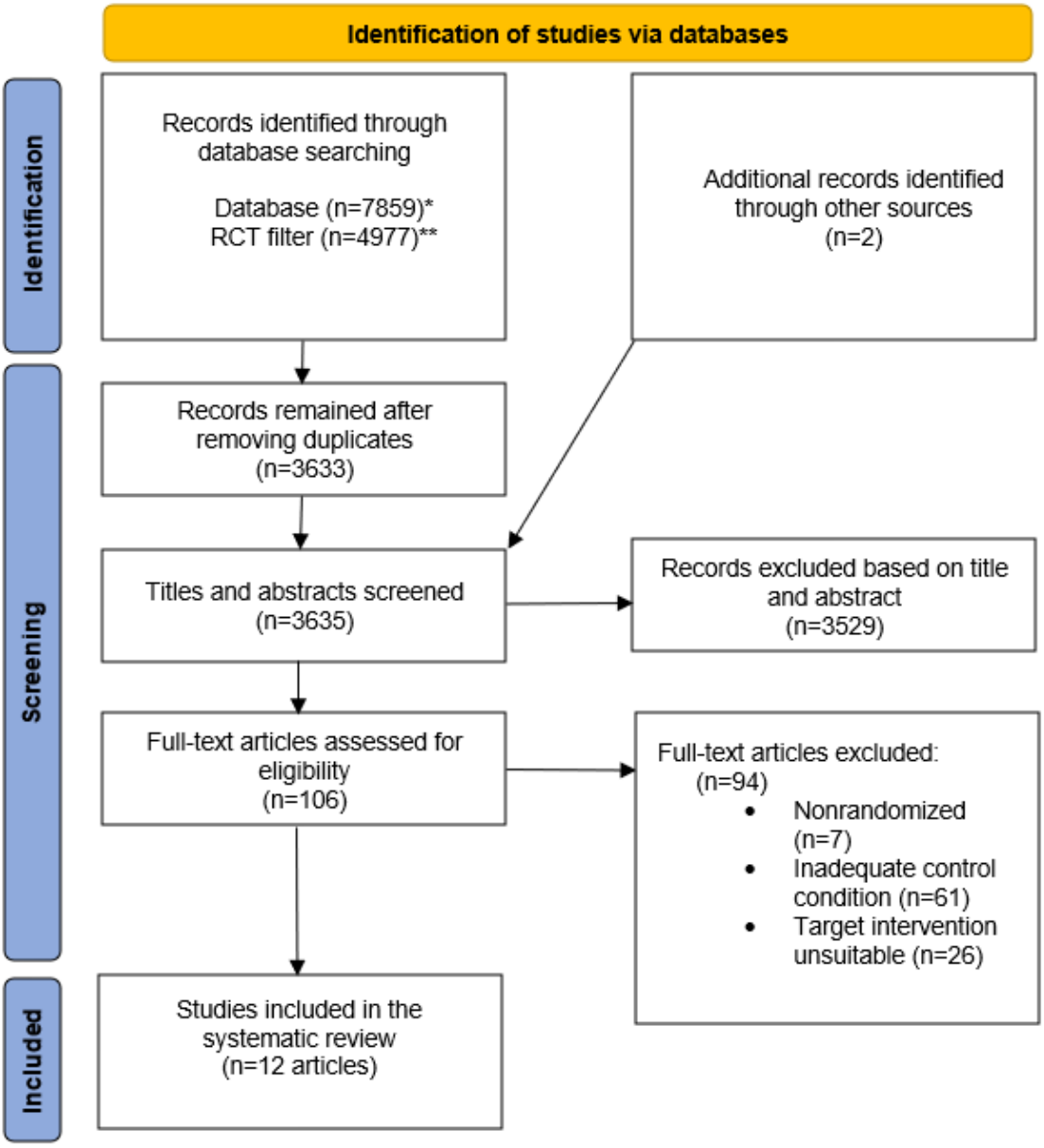


Table 1. Characteristics of included studies (N=754)[a].

Trial and articles	Trial sample size (% of N; F2F[b]; DD[c])	Youth age (y), mean (SD)	Male, n (%)	Feature	DI[d]	F2FI[e]	Sessions, mean (SD)
BFST-D[f]							
Duke et al [29], 2016	90 (11.9%; 46; 44)	15.02 (1.75)	55 (61.1)	HbA_{1c}[g] >9%	Videoconferencing	Clinic	5.8 (3.3)
Harris et al [58], 2015	90 (11.9%; 46; 44)	14.9[h] (1.7)	55 (61.1)[h]	HbA_{1c} >9%	Video conferencing	Clinic	5.8 (3.3)
Freeman et al [33], 2013	92 (12.2%; 45; 47)	15.1[i]	42 (45.7)[h]	HbA_{1c} >9%	Videoconferencing	Clinic	—[j]
Riley et al [57], 2015	82 (10.9%)[k]	14.1	55 (61.1)	HbA_{1c} >9%	Videoconferencing	Clinic	6.3 (3.4)
PAAS[l]							
Murry et al [56], 2019	421 (55.8%; 141; 141; CC[m]: 136)	—	195[h] (46)	PRB[n]	II[o]	GS[p]	—
Murry et al [59], 2019	412 (54.6%; 137; 138; CC: 137)	11.0	191[h] (46)	PRB	II	GS	—
Murry et al [30], 2018	412 (54.6%; 137; 138; CC: 137)	11.4	191 (46.4)	PRB	II	GS	4.0 (3.0)
F-PST[q]							
Kurowski et al [65], 2020	150 (19.9%; 34; 56; SG[r]: 60)	16.5 (1.1)	96 (64)	TBI[s]	Videoconferencing+OP[t,u]	Clinic	6.3 (2.6)
Wade et al [66], 2019	150 (19.9%; 34; 56; SG: 60)	16.5 (1.1)	96 (64)	TBI	Videoconferencing+OP[u]	Clinic	6.3 (2.6)
Wade et al [68],2019	150 (19.9%; 34; 56; SG: 60)	16.5 (1.1)	96 (64)	TBI	Videoconferencing+OP[u]	Clinic	6.3 (2.6)
Wade et al [67],2019	150 (19.9%; 34; 56; SG: 60)	15.5 (1.5)[v]	96 (64.0)	TBI	Videoconferencing+OP[u]	Clinic	6.3 (2.6)
SUCCEAT[w]							
Truttmann et al [64], 2020	102 (13.5%; 50; 52)	14.9 (1.9)	9 (8.9)	AN[x]	OP	GWS[y]	6.5 (2.0)

[a]Sum of n at the trial level. In case of inconsistencies in reported n across individual articles for each trial, the mode n was selected.

[b]F2F: face-to-face delivery.

[c]DD: therapist-guided digital delivery.

[d]DI: digital implementation.

[e]F2FI: face-to-face implementation.

[f]BFST-D: Behavioral Family Systems Therapy for Diabetes.

[g]HbA_{1C}: glycated hemoglobin.

[h]Calculated based on percentages and overall sample size provided.

[i]Calculated based on mean ages, sample sizes, and SDs provided.

[j]Not available.

[k]Only participants who completed the Child Depression Inventory measure (only those aged <18 y eligible).

[l]PAAS: Pathways for African American Success.

[m]CC: control condition.

[n]PRB: potential risk behavior.

[o]II: interactive interface.

[p]GS: group session.

[q]F-PST: Family Problem-Solving Therapy.

[r]SG: self-guided digital delivery.

[s]TBI: traumatic brain injury.

[t]OP: online program.

[u]Two distinct digital delivery conditions: (1) videoconferencing+web-based program and (2) online program.

[v]Most likely a clerical mistake in reporting.

[w]SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria.

[x]AN: anorexia nervosa.

[y]GWS: group workshop.

Characteristics of Included Articles

Number of Trials, Articles, and Sample Size

A total of 12 articles reporting on 4 trials were identified. One trial was reported on by a single article [64] (Supporting Carers of Children and Adolescents with Eating Disorders in Austria [SUCCEAT] trial), while the remaining 92% (11/12) of articles reported on a total of 3 trials. Of all 12 articles, 58% (n=7) of articles provided trial registration numbers, allowing for formally confirmed identification of the reported trial.

For the purposes of this study, the 4 trials are abbreviated according to their respective intervention names as follows: Behavioral Family Systems Therapy-Diabetes (BFST-D) trial [29,33,57,58], Family Problem-Solving Therapy (F-PST) trial [65-68], Pathways for African American Success (PAAS) trial [30,56,59], and the SUCCEAT trial [64].

As articles reporting on the BFST-D trial and articles reporting on the PAAS trial varied with regard to sample sizes (Table 1 and Multimedia Appendix 6), the total number of randomized participants across all trials could not be unambiguously determined. Given that for both trials, only 1 article each reported a different sample size as compared to all other articles reporting on the same trial, we reported the mode sample size for the total number of participants randomized to all conditions in all trials, including a control condition (N=754), and for the total number of participants randomized to face-to-face and digital delivery conditions (n=617). The same principle was applied for all reported participant characteristics, which varied in terms of proportions and absolute numbers depending on variation in reported sample sizes.

Participants

Some (4/12, 33%) articles reported on interventions delivered to youths with poorly controlled type 1 diabetes and their parents (BFST-D trial), 25% (n=3) articles reported on interventions delivered to parents, siblings, and youths identified as being at disproportionate risk for contracting HIV and sexually transmitted infections, teen parenthood, and substance use (PAAS trial). Moreover, 33% (4/12) articles reported on interventions delivered to youths with traumatic brain injury and their parents and other family members (F-PST trial), and 8% (1/12) articles reported on interventions delivered to parents of youths with anorexia nervosa (SUCCEAT trial). As there is inconsistent terminology between and within articles, the term *parent* is used here and throughout this paper to refer to parents, grandparents, legal guardians, and caregivers included in the respective articles. All identified interventions either involved families in intervention delivery (BFST-D, PAAS, and F-PST trials) or in terms of interpersonal target outcomes (SUCCEAT trial). Additional participant information is provided in Table 1 and Multimedia Appendix 7 [29,30,33,56-59,64-68]. Participant characteristics were reported to be similar across conditions in all articles, with the exception of the F-PST trial, which included a lower proportion of White participants in the face-to-face delivery condition, probably linked to quasi-randomization involving place of residence in the allocation procedure.

Face-to-Face and Digital Interventions and Delivery Format

Interventions were provided by health professionals in 75% (9/12) of articles (BFST-D, F-PST, and SUCCEAT trials) and by trained community leaders in 25% (3/12) articles (PAAS trial). Intervention content was similar across delivery modalities for all trials, with all articles explicitly mentioning the same intervention being delivered across conditions and all CPPs of interventions being delivered in both conditions (Table 1 and Multimedia Appendix 2).

All therapist-guided digital delivery was provided via videoconferencing software. Self-guided intervention delivery was provided via access to web-based modules for 33% (4/12) of the articles (F-PST trial) and for 25% (3/12) of the articles (PAAS trial) via an interactive interface visually representing users through avatars in a virtual environment. Face-to-face delivery was reported as individual in-clinic sessions for 67% (8/12) of the articles (BFST-D and F-PST trials), as group sessions at community centers for 25% (3/12) of the articles (PAAS trial), and in-clinic workshops for multiple individuals for 8% (1/12) of the articles (SUCCEAT trial).

Outcomes

Across all 12 articles, 56 outcome measures (ie, dependent variables) were reported, including cases of parent-report and youth-report measures of the same construct. Primary outcomes (8/56, 14%), 71% (40/56) secondary outcomes, and 14% (8/56) further, dropout, or attrition outcomes were identified. We have provided an overview of all identified outcomes in Multimedia Appendix 8 [29,33,56-59,64-68].

We summarized and structured data along three domains: (1) outcome type (ie, youth, parent, family functioning, and dropout, attrition, or further), (2) delivery modality type (face-to-face, therapist-guided digital, and self-guided digital), and (3) measurement time point (postintervention test and follow up test) to reduce clinical and methodological diversity within outcome data categories. We provide a more detailed overview of delivery modality types (eg, videoconferencing or in-clinic sessions) in Table 1, an overview of outcome measures (ie, dependent variables) in Multimedia Appendix 8, and measurement time points in Multimedia Appendix 6.

The application of these 3 domains to the categorization of data yielded 16 potential outcome categories: youth (Multimedia Appendix 9 [58,59,64-67]); parent (Multimedia Appendix 9); family functioning (Multimedia Appendix 10 [59]); and adherence, attrition, and further outcomes (Multimedia Appendix 11 [30,33,62,65,68]) for either therapist-guided digital versus face-to-face delivery (Multimedia Appendix 9) or self-guided digital versus face-to-face delivery modalities (Multimedia Appendices 9 and 10) at either posttest (Multimedia Appendices 9 and 10) or follow-up test (Multimedia Appendix 11) time points.

Risk of Bias

Overview

The 12 articles were rated as high risk for allocation concealment. For sequence generation, 75% (9/12) articles were rated as having some concerns. For randomization, 2 of the 4 trials used quasi-randomization. Some (9/12, 75%) articles were rated as having some concerns in terms of missing data. Overall proportions of RoB levels across articles are presented in Figure 2; individual RoB assessments at the article level are presented in Figure 3 [29,30,33,56-59,64-68]. Moreover, 92% (11/12) articles reported on 3 trials, with inconsistent reporting at times, which constituted a challenge for assessing the RoB, particularly for the domain of selective outcome reporting, as multiple articles reporting on the same trial were found to simultaneously overlap in some reported outcomes and diverge for other outcomes (eg, inconsistent reporting of the Diabetes Family Conflict Scale across articles reporting on the BFST-D trial).

Figure 2. Total proportions of articles' risk of bias assessments by assessment domain.

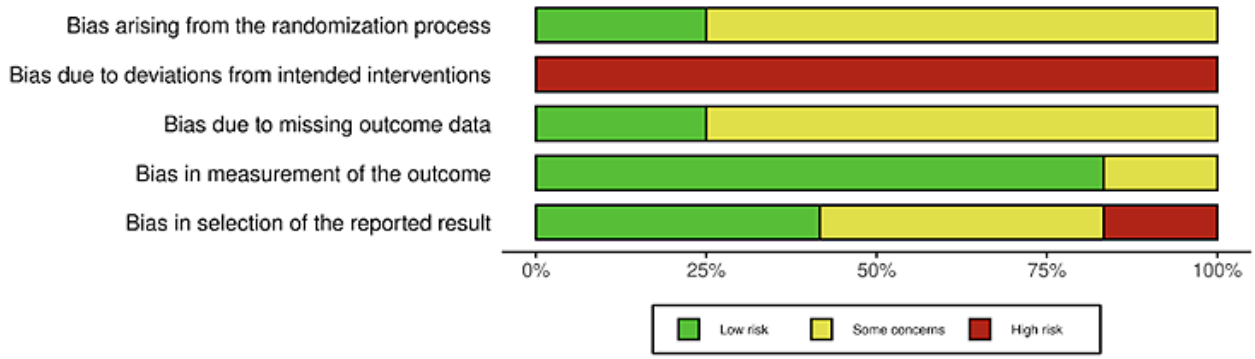
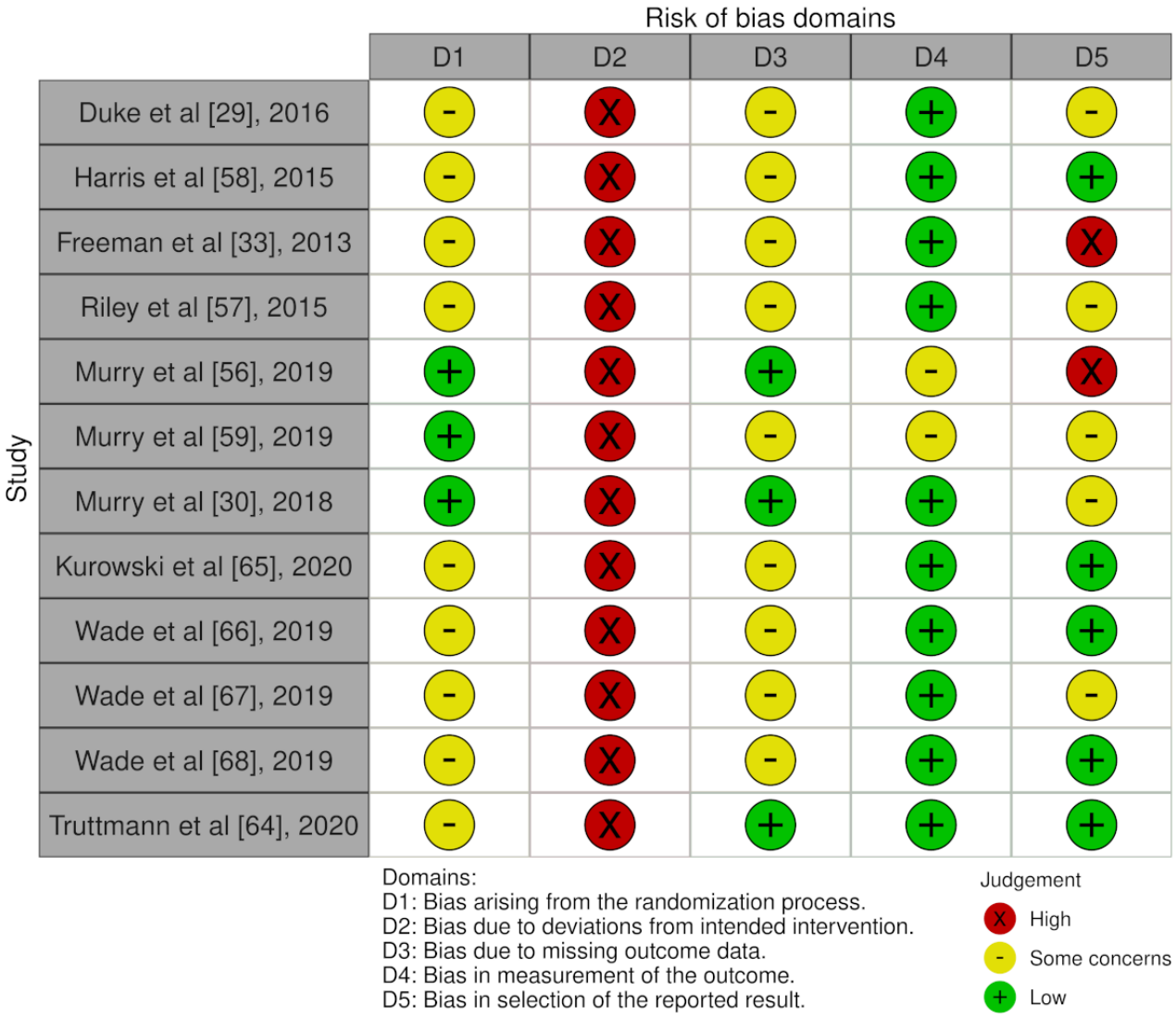


Figure 3. Risk of bias assessments of all included articles. D1: bias arising from the randomization process; D2: bias due to deviations from intended intervention; D3: bias due to missing outcome data; D4: bias in measurement of the outcome; D5: bias in selection of the reported result.



Sequence Generation

Of all 12 articles, 8% (n=1) reported a quasi-randomization temporal sequencing design (SUCCEAT trial); 33% (n=4) reported a quasi-randomized, multicenter RCT design (F-PST trial), in which participants were unequally allocated to the digital and face-to-face delivery conditions based on their place of residence (all participants who lived outside a 40-kilometer radius from the site were randomly allocated to the digital delivery conditions, and participants within the radius were randomly allocated to all conditions, favoring the face-to-face delivery condition by a ratio of 2:1); and 58% (n=7) reported randomized controlled trials (BFST-D and PAAS trials). Of those 7 articles, 57% (n=4) reported block randomization without providing information on allocation sequence concealment, increasing the RoB from sequence generation (BFST-D trial). Overall, 25% (3/12) of the included articles were rated low RoB arising from the randomization, while 75% (9/12) were rated as giving rise to some concerns.

Allocation Concealment

The question of adequate allocation blinding procedures (for participants and intervention providers) is a common concern in psychotherapy outcome research [71]. This also applies to the articles included in this review. Given the absence of a universally agreed-upon alternative approach [71], all articles were rated as high RoB in this domain.

Handling of Incomplete Data

For any of the 12 articles, not all outcome data were available. Some (2/12, 17%) of the articles provided evidence of data missing at random [56] or treated attrition as an outcome variable [30] and were rated as having low RoB. Most (9/12, 75%) articles did not adequately analyze if missing outcome data did not potentially bias the results and were rated as giving rise to some concerns.

Outcome Assessment

RoB arising from the measurement of the outcome variables was rated *low risk* for 83% (10/12) of the articles and with *some concerns* for 17% (2/12) of the articles that used a large proportion of outcome measures for which judgments regarding validity and reliability could not be made [56,59] (Multimedia Appendix 8).

Selective Reporting

As indicated earlier, RoB from selective reporting was challenging to assess given the interaction of multiple articles reporting on the same trial and inconsistencies across articles on the same trials. This was further exacerbated in the case of a trial for which no preregistration could be identified (PAAS trial) and of an article exclusively reporting on unregistered outcomes of a preregistered trial [33]. Some (5/12, 42%) articles were rated as giving rise to some concerns, 9% (5/12) were rated as having low RoB, and 17% (2/12) of the articles were rated as having high RoB.

Appropriateness of a Meta-Analysis to Determine the Efficacy of Delivery Modalities, Adherence, and Attrition

Meta-analytic aggregation was deemed inappropriate for all delivery modality comparisons of youth, parent, and family functioning outcomes due to the relatively low number of trials within the respective outcome categories, combined with the presence of substantial clinical and methodological diversity across trials [63]. Meta-analytic aggregation of the number of sessions attended and attrition was conducted because heterogeneity was lower and the number of eligible trials was higher for these outcomes as compared to the other outcome measures reported and included (Figure 4 [58,64,65] and Figure 5 [58,64,65]). The substantial clinical and methodological diversity across trials remains problematic for drawing meaningful conclusions using meta-analytic aggregations. However, a preliminary quantitative synthesis was included as it still serves as a starting point for discussion and future research.

Figure 4. Meta-analysis of number of sessions attended in face-to-face versus digital delivery modality comparison.

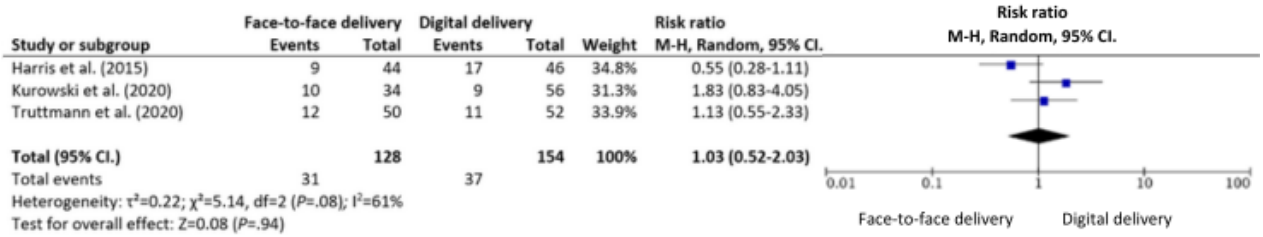
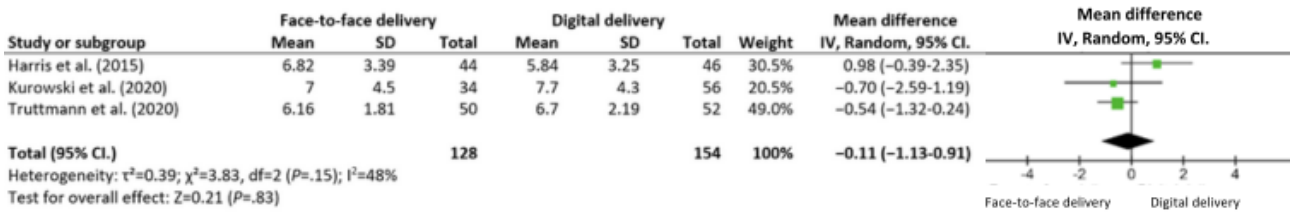


Figure 5. Meta-analysis of drop-out (event) risk ratio in face-to-face versus digital delivery modality comparison.



Nonaggregated Analyses for Face-to-Face and Digital Delivery Modalities

Of all 56 reported outcomes (ie, dependent variables) across all trials and articles, 25% (n=14) outcomes from 2 trials (13/14, 93% in the F-PST trial and 1/14, 7% in the SUCCEAT trial) showed MDs with corresponding CIs falling either (1) fully within our predefined range of minimal important difference of –0.20 to 0.20 or (2) fully outside of this range. These outcomes are reported in Table 2 and Multimedia Appendix 12. All remaining 75% (42/56) outcomes showed MDs with corresponding CIs partially located within and partially located outside this range. Among these, 4% (2/56) further outcomes, 2% (1/56) family functioning outcome, and 2% (1/56) youth outcome were reported as substantially different between delivery modalities in the original articles while still including our range for the minimal important difference. Multimedia Appendix 9 shows a full list of all MDs and CIs of outcomes.

Table 2. Outcomes with substantial differences (95% CI excluding 0) or equivalence across all trials (N=252; sum of n at the trial level).

Outcome type, postintervention test and follow-up test, trial, and article	Outcome	Type of DD[a]	F2F[b] delivery, n	F2F delivery, mean (SD)	DD, n	DD, mean (SD)	Mean difference (95% CI)	Directionality
Youth								
Postintervention test								
F-PST[c]; Wade et al [67], 2019								
	PedsQL[d], parent report	TG[e]	22	62 (15.71)	43	70.9 (18.49)	–8.90 (–17.48 to –0.32)	Favors DD
	PedsQL, parent report	SG[f]	22	62 (15.71)	51	71.7 (18.85)	–9.70 (–18.06 to –1.34)	Favors DD
Follow-up test								
F-PST; Wade et al [67], 2019								
	HBI[g] somatic, youth report	TG	22	9.8 (4.92)	44	6.6 (5.90)	3.20 (0.50 to 5.90)	Favors DD
	HBI cognitive, parent report	SG	22	19.1 (8.58)	48	12.3 (9.91)	6.80 (2.25 to 11.35)	Favors DD
	HBI somatic, youth report	SG	22	9.8 (4.92)	48	6.8 (5.89)	3.00 (0.35 to 5.65)	Favors DD
	PedsQL, parent report	SG	22	67 (16.18)	48	76.6 (18.98)	–9.60 (–18.23 to –0.97)	Favors DD
Parent								
Postintervention test								
F-PST; Wade et al [66], 2019								
	CES-D[h]	TG	22	15.1 (1.81)	43	12.8 (1.53)	2.30 (1.42 to 3.18)	Favors DD
	BSI[i]	TG	22	56.1 (2.46)	43	53.3 (2.02)	2.80 (1.61 to 3.99)	Favors DD
	CES-D	SG	22	15.1 (1.81)	51	13.5 (1.42)	1.60 (0.75 to 2.45)	Favors DD
	BSI	SG	22	56.1 (2.46)	51	54.7 (1.88)	1.40 (0.25 to 2.55)	Favors DD
SUCCEAT[j]; Truttmann et al [64], 2020								
	SCL[k]-90-R	TG	48	0.24 (0.26)	46	0.31 (0.36)	–0.07 (–0.20 to 0.06)	Equivalence
Follow-up test								
F-PST; Wade et al [66], 2019								
	BSI	TG	22	51.7 (2.42)	44	53.3 (2.02)	–1.60 (–2.77 to –0.43)	Favors F2F delivery
	CES-D	SG	22	11.8 (2.03)	48	13.2 (1.58)	−1.40 (−2.36 to −0.44)	Favors F2F delivery
Wade et al [68], 2019								
	P-PE[l]	TG	22	9.0 (1.4)	43	7.8 (2.1)	1.20 (0.34 to 2.06)	Favors F2F delivery

[a]DD: digital delivery.

[b]F2F: face-to-face.

[c]F-PST: Family Problem-Solving Therapy.

[d]PedsQL: Pediatric Quality of Life Inventory.

[e]TG: therapist guided.

[f]SG: self-guided.

[g]HBI: Health and Behaviour Inventory.

[h]CES-D: Center for Epidemiological Studies Depression Scale.

[i]BSI: Brief Symptom Inventory.

[j]SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria.

[k]SCL: Symptom Checklist 90-Revised Global Severity Index.

[l]P-PE: Parent Rated Program Evaluation.

Of the 14 outcomes falling fully within or fully outside the minimal important difference range, superiority of the digital delivery condition was found for 71% (n=10) outcomes. Of these 10 outcomes, 60% (n=6) were youth outcomes and 40% (n=4) were parent outcomes. The quality of life of youths, reported by parents (Pediatric Quality of Life Inventory, parent report) at post intervention test, was substantially higher for both therapist-guided and self-guided digital delivery modalities compared to face-to-face delivery modalities and remained substantially higher for the self-guided digital delivery condition at follow-up. Somatic symptoms of youths, reported by them (Health and Behaviour Inventory somatic, youth report) at follow-up, were substantially higher in the face-to-face delivery condition compared to both therapist-guided and self-guided digital delivery conditions. In addition, cognitive symptoms of youths, reported by parents (Health and Behaviour Inventory cognitive, parent report) at follow-up, were substantially higher in the face-to-face delivery condition compared to the self-guided digital delivery condition. Depressive symptoms of parents (Center for Epidemiological Studies Depression Scale) at post intervention test were substantially higher in the face-to-face delivery condition compared to both therapist-guided and self-guided digital delivery conditions. Similarly, mental symptoms of parents (Brief Symptom Inventor) at post intervention measurement were substantially higher in the face-to-face delivery condition compared to both therapist-guided and self-guided digital delivery conditions.

Superiority of the face-to-face delivery condition was found for 3 outcomes. All of these outcomes were parent outcomes from the F-PST trial at follow-up. In contrast to comparisons at post intervention measurement, depressive symptoms of parents (Center for Epidemiological Studies Depression Scale) at follow-up were substantially higher in the self-guided digital delivery condition compared to the face-to-face delivery condition. Similarly, contrasting with comparisons at post intervention measurement, mental symptoms of parents (Brief Symptom Inventory) at follow-up were substantially higher in the therapist-guided digital delivery condition compared to the face-to-face delivery condition. Program evaluation by parents was more favorable in the face-to-face delivery condition compared to the therapist-guided digital delivery condition.

Only 2% (1/56) of outcomes fell fully within the minimal important difference range. In the SUCCEAT trial, psychopathological symptoms of parents (Symptom Checklist 90-Revised Global Severity Index) at post intervention test were equal between therapist-guided digital and face-to-face delivery conditions.

Meta-Analytic Aggregation of Adherence and Attrition

A total of 9 articles reporting on all 4 trials (BFST-D trial [29,57,58], PAAS trial [59], F-PST trial, and SUCCEAT trial) reported on the number of sessions attended, and all (11/12, 92%) articles, except for 1 (PAAS trial [57]), reporting on all trials reported on attrition. The number of sessions attended and attrition were meta-analyzed, and the meta-analytic aggregations are reported in Figures 4 and 5. The heterogeneity between trials was moderate (I^2=48% for number of sessions attended) to high (I^2=61% for attrition), and given the low number of included trials, this heterogeneity estimate might be inaccurate [72], and conclusions from the meta-analyses should thus be interpreted with caution.

Discussion

Principal Findings

We identified 12 articles reporting on 4 trials from which data were extracted (Multimedia Appendix 5). Of note, despite broad inclusion criteria for participants, participants included only youths or parents of youths. Recipients of interventions were youths with poor diabetic control, traumatic brain injuries, increased risk behavior likelihood, and parents of youths with anorexia nervosa. A total of 56 relevant outcomes were identified. Two trials provided digital intervention delivery via videoconferencing, one via an interactive graphic interface, and the other via a web-based program. RoB judgments varied substantially across domains and articles, with largely *some concerns* judgments for randomization, missing outcome data, and selective reporting. Applying a minimal important difference criterion to indicate equivalence or superiority, nonaggregated analyses of MDs and CIs between delivery modalities yielded varying types of comparative outcomes across the included studies: superiority of the digital delivery modality and superiority of the face-to-face delivery modality (CIs falling fully outside the range of the minimal important difference) and equivalence between delivery modalities (CIs falling fully inside the range of the minimal important difference), with most comparative outcomes, indicating neither superiority of one modality nor equivalence between delivery modalities (CIs falling within and outside the range of the minimal important difference). Overall, our analyses did not reveal any strong evidence for the superiority of face-to-face compared to digital delivery conditions and might tentatively instead indicate a higher proportion of favorable effects of digital delivery modalities for certain outcomes and certain contexts. Due to the qualitative nature of this analysis, the limited number of studies included in this review, and the large heterogeneity between studies, results need to be interpreted with caution. In addition, more research is needed to draw meaningful conclusions on the comparative efficacy of digital versus face-to-face delivery modalities of ST interventions.

We found a limited and insufficient amount of evidence to meta-analytically estimate the comparative efficacy of digital

versus face-to-face delivery modalities of ST interventions with regard to most psychological, somatic, and behavioral parent and youth outcomes, as well as family functioning and further outcomes at either postintervention test or follow-up test. Meta-analytic analysis was deemed appropriate only for attrition and the number of sessions attended. We compared digital versus face-to-face delivery modalities for attrition and the number of sessions attended, showing no evidence for the superiority of either delivery modality in these domains.

Comparison With Prior Work

Our findings are largely consistent with related findings from a recent meta-analysis on family therapy, including 3 trials reporting no evidence for a difference at postintervention test between therapist-guided digital versus face-to-face delivery of interventions for youth outcomes and parent outcomes [6]. Similarly, our findings are consistent with findings from a recent meta-analysis on therapist-guided digital parenting interventions that included 4 trials reporting no evidence for a difference at posttest for youth outcomes between therapist-guided digital and face-to-face delivery conditions [7]. None of this prior work, including ours, either strongly supports the superiority of face-to-face over digital delivery conditions or vice versa.

This study adds to research as the first to contribute to the existing body of research from an ST standpoint comparing RCTs of digital and face-to-face delivery modalities of the same intervention. While showing considerable overlap with family therapy, couple therapy, and parenting interventions, ST interventions are characterized by a largely independent theoretical framework and set of therapeutic techniques. Recent reviews in related areas of research [6,7,10,47,73] include interventions with different conceptual backgrounds (eg, family-based cognitive behavioral therapy and the integrated family intervention for child conduct problems [6]) under umbrella terms, such as family therapy or parenting interventions. The theory-driven approach we adopted for this systematic review helps to increase granularity by only including interventions based on ST content and conceptual background.

Strengths and Limitations

Our systematic review has several strengths. First, we applied rigorous methodology, which was preregistered and followed PRISMA guidelines [26]. Second, we applied a search strategy designed for high sensitivity to 6 databases, with an extensive range of complimentary searches coupled with consistent application of inclusion criteria on study design, intervention type, and delivery modalities, which enabled the descriptive synthesis of published evidence based on diverse target populations, different contexts, and different digital delivery modalities. Third, including ST interventions by identifying CPPs and testing them against our definition criteria ensured high levels of reproducibility and validity.

However, there are also limitations to be considered when interpreting the findings of this study. First, the definition of ST applied here is an adapted version of one of several, at times inconsistent, definitions of ST currently used [74,75]. However, our definition is based on the widely accepted definition by von Sydow et al [19], which has also been used as the basis for large-scale public policy and insurance regulations. Second, the operationalized definition criteria were only applied to identified CPPs of the interventions and for criterion 1 (namely, “perceives behavior and mental symptoms within the context of the social systems people live in”), conceptual background to make inclusion or exclusion judgments during screening. This approach may have excluded additional potentially relevant information about the interventions not provided as part of the description of the intervention itself (eg, the information provided in cited articles). However, we judged this approach to be the most beneficial to this study’s replicability by reducing the arbitrary selection of intervention features as much as possible. Third, given that this study was, to our knowledge, the first to investigate digital versus face-to-face delivery modalities for ST interventions, no a priori categorical constraints were defined with regard to participants, outcomes, or contexts. Instead, constraints were defined post hoc, with the goal of maximizing diversity between and minimizing diversity within subgroups. Of note, despite a few constraints regarding the type of participant population, the sample from which participants were recruited was found to be comparatively narrowly defined (youths or parents of youths). Fourth, while our definition of the minimal important difference is in line with current guidelines [19] and current research in related fields [69], also based on Cohen *d* cutoff of 0.2 for a small effect, there are other approaches, such as anchor-based estimates [76], which might yield different ranges for the minimal important differences for the scales included in this study. However, given the fact that the required data for other approaches toward defining the minimally important differences, such as anchors or expert ratings, are currently not available for all identified outcome measures, our approach based on a Cohen *d* cut-off is acceptable and feasible.

Generalizability

Some factors limit the generalizability of our findings. First, while the identification of a low number of trials constitutes a finding in its own right, it limits the appropriateness of meta-analytic aggregation. This was exacerbated by the fact that multiple articles with potentially relevant findings reporting on the same trial were identified and that some articles, in some cases, showed inconsistencies in reporting data (eg, on sample size). Second, despite post hoc introduction of categories, diversity within these categories remained substantial (eg, in terms of participants’ race, differences in face-to-face delivery formats, or different outcome domains), further limiting the appropriateness of meta-analytic aggregation. The differential impact of the type of digital modality used in each trial further limits the results of our study, and due to the low number of trials, it was not possible to analyze these tools separately. Third, the high diversity between trials (eg, in terms of delivery modalities, patient features, outcomes, and contexts) combined with the low overall number of trials limits the generalizability of this study’s findings to other populations and delivery contexts. Fourth, the low number of identified trials did not allow for subgroup analyses based on RoB judgments. Fifth, 3 of the 4 trials were conducted in the United States, and 3 of the 4 trials recruited a substantial majority of White participants, further limiting generalizability to other contexts and target

populations. Sixth, the influence of factors such as paraverbal communication, subject of discussion, and client-practitioner dynamics is crucial for understanding psychotherapeutic processes. Although RCTs are seen as the gold standard in evidence-based medicine, they may not adequately represent these factors, and the same applies to our study by extension.

Clinical and Research Implications

This review has some important clinical and research implications. Before digital interventions can be accepted unanimously and unambiguously, solid evidence is needed to be built on specific criteria for evaluating the appropriateness of digital modes of delivering ST interventions. Digital delivery is rapidly increasing despite insufficient evidence on the comparative efficacy of digital and face-to-face delivery modalities of ST. This might explain some of the reported clinicians' resistance to implementing digital interventions in their practices [77] and should caution against replacing face-to-face with digital delivery modalities without careful consideration. At the same time, given the evidence base supporting the efficacy of digitally delivered systemic psychotherapy (though currently still not as broad as that for face-to-face delivery), the largely inconclusive evidence regarding the comparative efficacy of delivery modalities paired with some indicators for the superiority of digital delivery modalities of ST for certain populations and contexts could also be interpreted as a caution against adopting face-to-face delivery of ST interventions as the clinical default. Our study might serve as a first tentative indicator that different populations benefit differently from different delivery modalities. One could, for instance, look at differences between significant outcomes from the included trials in our study and speculate if parents might be more likely to benefit from face-to-face delivery than youths who, in turn, might be more likely to benefit from digital delivery modalities. Potential avenues for future research in this regard could, for instance, investigate the impact of the parents' role as facilitators of self-guided group interventions. At this point, these points of departure for future research remain speculative, given the current lack of clear evidence regarding the comparative efficacy of digital and face-to-face delivery modalities of ST. To that end, more studies with equivalence design are required.

According to Greene et al [78], equivalence studies should include 4 key aspects in their design: an active comparator condition, identifying appropriate equivalence margins, calculating adequate sample size estimations, and doing appropriate statistical analysis [78]. Our review highlights the need for research implementing all these key aspects. Furthermore, future research should explore subgroups or conduct sensitivity analyses to understand the limitations and the differential advantages of digital and face-to-face delivery modalities. The questions, such as how different types of digital interventions compare to face-to-face interventions and which intervention is efficacious for target populations, specific contexts, and various modes of digital interventions, including therapist-guided, self-guided, videoconferencing, virtual augmented reality, and smartphone apps, remain unanswered. Moreover, research on digital apps and interventions that, even if not directly involved in therapy, may provide several benefits in supporting those delivering and receiving mental health interventions is required. There is a need to investigate digital modes of delivery that could potentially complement face-to-face interventions through blended forms of care (refer to the study by Erbe et al [79]), specifically in the context of ST. Overall, more high-quality controlled trials using similar outcome measures as well as systematic reviews and meta-analyses are required to scrutinize under what circumstances digital ST interventions work best, especially given the increasing amount of available modes of delivery and the increase in digital delivery of mental health care in general.

Conclusions

Taken together, our systematic review indicated that the current evidence regarding the comparative efficacy of digital and face-to-face delivery modalities of ST interventions does not allow conclusions to be drawn regarding equivalence or superiority for youth, parent, family functioning, adherence, attrition, or further outcomes. There is a dearth of RCTs comparing the efficacy of comparable ST interventions delivered via both digital and face-to-face delivery modalities. The low number of trials combined with high clinical and methodological diversity between trials did not permit drawing meaningful conclusions from meta-analytic aggregation, and neither equivalence nor superiority of one modality can be excluded at this point. Nonaggregated analyses of MDs and CIs between delivery modalities indicated that neither equivalence nor superiority of either modality can be excluded at this stage for most outcomes, target populations, and contexts. As digital delivery of ST interventions is becoming increasingly ubiquitous, more research is urgently required to investigate how digital delivery of ST interventions compares to face-to-face delivery of the same interventions. There is a need for more research with a particular focus on different types of digital delivery modalities, including blended forms of mental health care, investigating potential variation in efficacy across different settings, populations, outcomes, and contexts.

Acknowledgments

GM received funding from the Stanley Thomas Johnson Stiftung and Gottfried und Julia Bangerter-Rhyner-Stiftung (PC 28/17 and PC 05/18), from Gesundheitsförderung Schweiz (18.191/K50001), from the Swiss Heart Foundation (FF21101), from the Research Foundation of the International Psychoanalytic University Berlin (5087), from the Swiss National Science Foundation (100014_135328), from the Hasler Foundation (23004), in the context of a Horizon Europe project from the Swiss State Secretariat for Education, Research and Innovation contract 22.00094, and from Wings Health in the context of a proof-of-concept study. GM and LK received funding from the German Federal Ministry of Education and Research (budget item 68606). GM and MB

received funding from the Research Foundation of the International Psychoanalytic University Berlin (5217). The article was funded and supported by the Open Access Fund of Universität Trier and by the German Research Foundation (DFG).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request. A list of articles excluded with respective justifications for exclusions at the full-text screening stage can be made available upon request.

Authors' Contributions

PE and MB conceptualized the study, conducted data analysis and interpretation, and wrote the manuscript draft. PE, MB, and JB contributed to data collection. PE, MB, GM, LK, and JB revised the manuscript. PE, MB, GM, LK, and JB contributed to the final approval of the manuscript for publication.

Conflicts of Interest

GM is the cofounder and holds stock in Therayou AG, which is active in the field of digital and blended mental health care. GM receives royalties from publishing companies as an author, including a book published by Springer, and an honorarium from Lundbeck for speaking at a symposium. Furthermore, GM is compensated for providing psychotherapy to patients, acting as a supervisor, serving as a self-experience facilitator (“Selbsterfahrungsleiter”), and for postgraduate training of psychotherapists, psychosomatic specialists, and supervisors. Other authors do not declare any conflicts.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 34 KB - ijmr_v14i1e46441_app1.docx]

Multimedia Appendix 2
Intervention inclusion criterion and constituent primary parts of interventions meeting definition criteria of systemic psychotherapy.
[DOCX File , 25 KB - ijmr_v14i1e46441_app2.docx]

Multimedia Appendix 3
Search terms, structure of search string, and database adaptations.
[DOCX File , 34 KB - ijmr_v14i1e46441_app3.docx]

Multimedia Appendix 4
Data extracted.
[DOCX File , 19 KB - ijmr_v14i1e46441_app4.docx]

Multimedia Appendix 5
Reported statistics on face-to-face versus digital delivery.
[DOCX File , 39 KB - ijmr_v14i1e46441_app5.docx]

Multimedia Appendix 6
Characteristics of included articles.
[DOCX File , 52 KB - ijmr_v14i1e46441_app6.docx]

Multimedia Appendix 7
Outcome measures overview.
[DOCX File , 23 KB - ijmr_v14i1e46441_app7.docx]

Multimedia Appendix 8
Comparison tables for outcomes.
[DOCX File , 44 KB - ijmr_v14i1e46441_app8.docx]

Multimedia Appendix 9
Family functioning outcomes.
[DOCX File , 54 KB - ijmr_v14i1e46441_app9.docx]

Multimedia Appendix 10
Adherence, attrition, and further outcomes.
[DOCX File , 20 KB - ijmr_v14i1e46441_app10.docx]

Multimedia Appendix 11
Screened reviews and meta-analyses.
[DOCX File , 33 KB - ijmr_v14i1e46441_app11.docx]

Multimedia Appendix 12
Flowchart of participant characteristics, measurement time points, intervention delivery type, and superiority or equivalence of delivery modalities. BSI: Brief Symptom Inventory; CES-D: Center for Epidemiological Studies Depression Scale; DD: digital delivery; F2F: face-to-face; F-PST: Family Problem-Solving Therapy; HBI: Health and Behaviour Inventory; P-PE: Parent Rated Program Evaluation; SCL: Symptom Checklist 90-Revised Global Severity Index; SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria; PedsQL: Pediatric Quality of Life Inventory.
[PNG File , 1130 KB - ijmr_v14i1e46441_app12.png]

References

1. Dack C, Pal K, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
2. Pompili S, Orsolini L, Salvi V, Volpe U. A systematic review on TeleMental health in youth mental health: focus on anxiety, depression and obsessive-compulsive disorder. Medicina (Kaunas) 2021 Jul 31;57(8):793 [FREE Full text] [doi: 10.3390/medicina57080793] [Medline: 34440999]
3. Doss BD, Knopp K, Roddy MK, Rothman K, Hatch SG, Rhoades GK. Online programs improve relationship functioning for distressed low-income couples: results from a nationwide randomized controlled trial. J Consult Clin Psychol 2020 Apr;88(4):283-294. [doi: 10.1037/ccp0000479] [Medline: 32134290]
4. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
5. Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the COVID-19 pandemic. Riv Psichiatr 2020;55(3):137-144. [doi: 10.1708/3382.33569] [Medline: 32489190]
6. McLean SA, Booth AT, Schnabel A, Wright BJ, Painter FL, McIntosh JE. Exploring the efficacy of telehealth for family therapy through systematic, meta-analytic, and qualitative evidence. Clin Child Fam Psychol Rev 2021 Jun;24(2):244-266 [FREE Full text] [doi: 10.1007/s10567-020-00340-2] [Medline: 33492545]
7. Florean IS, Dobrean A, Păsărelu CR, Georgescu RD, Milea I. The efficacy of internet-based parenting programs for children and adolescents with behavior problems: a meta-analysis of randomized clinical trials. Clin Child Fam Psychol Rev 2020 Dec 08;23(4):510-528. [doi: 10.1007/s10567-020-00326-0] [Medline: 32897527]
8. Baumel A, Correll CU, Birnbaum M. Adaptation of a peer based online emotional support program as an adjunct to treatment for people with schizophrenia-spectrum disorders. Internet Interv 2016 May;4:35-42 [FREE Full text] [doi: 10.1016/j.invent.2016.03.003] [Medline: 30135788]
9. Carr A. Couple therapy, family therapy and systemic interventions for adult‐focused problems: the current evidence base. J Fam Ther 2018 Jun 13;41(4):492-536. [doi: 10.1111/1467-6427.12225]
10. de Boer K, Muir SD, Silva SS, Nedeljkovic M, Seabrook E, Thomas N, et al. Videoconferencing psychotherapy for couples and families: a systematic review. J Marital Fam Ther 2021 Apr 10;47(2):259-288. [doi: 10.1111/jmft.12518] [Medline: 33837968]
11. Köhnen M, Dreier M, Seeralan T, Kriston L, Härter M, Baumeister H, et al. Evidence on technology-based psychological interventions in diagnosed depression: systematic review. JMIR Ment Health 2021 Feb 10;8(2):e21700 [FREE Full text] [doi: 10.2196/21700] [Medline: 33565981]
12. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Lingley-Pottie P, et al. Internet-assisted parent training intervention for disruptive behavior in 4-year-old children: a randomized clinical trial. JAMA Psychiatry 2016 Apr 01;73(4):378-387. [doi: 10.1001/jamapsychiatry.2015.3411] [Medline: 26913614]
13. von Sydow K, Retzlaff R, Beher S, Haun MW, Schweitzer J. The efficacy of systemic therapy for childhood and adolescent externalizing disorders: a systematic review of 47 RCT. Fam Process 2013 Dec 11;52(4):576-618. [doi: 10.1111/famp.12047] [Medline: 24102196]
14. von Sydow K, Retzlaff R. Aktueller stand der systemischen therapie. Psychotherapeut 2021 Oct 28;66(6):469-477. [doi: 10.1007/S00278-021-00547-W]
15. Retzlaff R, von Sydow K, Beher S, Haun MW, Schweitzer J. The efficacy of systemic therapy for internalizing and other disorders of childhood and adolescence: a systematic review of 38 randomized trials. Fam Process 2013 Dec 05;52(4):619-652. [doi: 10.1111/famp.12041] [Medline: 24329407]

16. Riedinger V, Pinquart M, Teubert D. Effects of systemic therapy on mental health of children and adolescents: a meta-analysis. J Clin Child Adolesc Psychol 2017 Oct 14;46(6):880-894. [doi: 10.1080/15374416.2015.1063427] [Medline: 26467300]
17. Stratton P, Lask J. The development of systemic family therapy for changing times in the United Kingdom. Contemp Fam Ther 2013 Mar 13;35(2):257-274. [doi: 10.1007/S10591-013-9252-8]
18. von Sydow K, Stefan Beher S, Retzlaff R, Schweitzer J. Wie wirksam ist Systemische Psychotherapie/Familientherapie? Psychiatr Prax 2008 Feb;8(1):20-27. [doi: 10.1026/1860-7357.8.1.20]
19. Allgemeine Methoden. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). URL: https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf [accessed 2024-04-29]
20. Borcsa M, Pomini V. La diffusione delle pratiche digitali nella psicoterapia sistemica: lo stato dell'arte in Europa e oltre. In: Manfrida G, Albertini V, Eisenberg E, editors. La clinica e il web. Risorse tecnologiche e comunicazione psicoterepeutica online. Milano, Italy: FrancoAngeli; Jul 07, 2020:37-60 URL: https://www.stateofmind.it/2020/09/clinica-web-recensione/
21. Lebow JL. Couple and family therapy (and families) in the digital age. Fam Process 2021 Sep 27;60(3):689-693. [doi: 10.1111/famp.12705] [Medline: 34580874]
22. Probst T, Stippl P, Pieh C. Changes in provision of psychotherapy in the early weeks of the COVID-19 lockdown in Austria. Int J Environ Res Public Health 2020 May 27;17(11):3815 [FREE Full text] [doi: 10.3390/ijerph17113815] [Medline: 32471295]
23. Burbach FA, Pote H. Digital approaches – a paradigm shift? J Fam Ther 2021 Mar 18;43(2):169-184. [doi: 10.1111/1467-6427.12336]
24. Borcsa M, Pomini V, Saint‐Mont U. Digital systemic practices in Europe: a survey before the COVID‐19 pandemic. J Fam Ther 2020 Sep 09;43(1):4-26. [doi: 10.1111/1467-6427.12308]
25. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Pinsof W. Integrative systemische therapie (IST). In: von Sydow KB, Borst U, editors. Systemische Therapie in der Praxis. Berlin, Germany: Beltz, Weinheim; Nov 13, 2018:963-975.
28. von Sydow KB, Borst U. Systemische Therapie in der Praxis. Berlin, Germany: Beltz; Nov 13, 2018.
29. Duke DC, Wagner DV, Ulrich J, Freeman KA, Harris MA. Videoconferencing for teens with diabetes: family matters. J Diabetes Sci Technol 2016 Jul;10(4):816-823 [FREE Full text] [doi: 10.1177/1932296816642577] [Medline: 27075708]
30. Murry VM, Berkel C, Liu N. The closing digital divide: delivery modality and family attendance in the pathways for African American success (PAAS) program. Prev Sci 2018 Jul 22;19(5):642-651 [FREE Full text] [doi: 10.1007/s11121-018-0863-z] [Medline: 29356928]
31. Kurowski BG, Stancin T, Taylor HG, McNally KA, Kirkwood MW, Cassedy A, et al. Comparative effectiveness of family problem-solving therapy (F-PST) for adolescents after traumatic brain injury: Protocol for a randomized, multicenter, clinical trial. Contemp Clin Trials Commun 2018 Jun;10:111-120 [FREE Full text] [doi: 10.1016/j.conctc.2018.04.001] [Medline: 30023445]
32. Franta C, Philipp J, Waldherr K, Truttmann S, Merl E, Schöfbeck G, et al. Supporting carers of children and adolescents with eating disorders in Austria (SUCCEAT): study protocol for a randomised controlled trial. Eur Eat Disord Rev 2018 Sep 06;26(5):447-461 [FREE Full text] [doi: 10.1002/erv.2600] [Medline: 29732651]
33. Freeman KA, Duke DC, Harris MA. Behavioral health care for adolescents with poorly controlled diabetes via Skype: does working alliance remain intact? J Diabetes Sci Technol 2013 May 01;7(3):727-735 [FREE Full text] [doi: 10.1177/193229681300700318] [Medline: 23759406]
34. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc 2018 Oct 04;106(4):531-541 [FREE Full text] [doi: 10.5195/jmla.2018.283] [Medline: 30271302]
35. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res Synth Methods 2018 Dec 07;9(4):602-614 [FREE Full text] [doi: 10.1002/jrsm.1287] [Medline: 29314757]
36. Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, et al. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. J Clin Epidemiol 2021 May;133:140-151 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.11.003] [Medline: 33171275]
37. Nussbaumer-Streit B, Mayr V, Dobrescu A, Chapman A, Persad E, Klerings I, et al. The effectiveness of quarantine alone or in combination with other public health measures to control coronavirus disease 2019: a rapid review. Cochrane Database Syst Rev 2020 Mar 27;9(Suppl 1):1 [FREE Full text] [doi: 10.1002/14651858.CD202001]
38. Hefti S, Pérez T, Fürstenau U, Rhiner B, Swenson CC, Schmid M. Multisystemic therapy for child abuse and neglect: do parents show improvement in parental mental health problems and parental stress? J Marital Fam Ther 2020 Jan 05;46(1):95-109. [doi: 10.1111/jmft.12367] [Medline: 30516844]

39. Tomm K. Introducing the IPscope: a systemic assessment tool for distinguishing interpersonal patterns. In: Tomm K, St. George S, Wulff D, Strong T, editors. Patterns in Interpersonal Interactions: Inviting Relational Understandings for Therapeutic Change. New York, NY: Routledge; 2014:13-35.
40. Ataoglu A, Ozcetin A, Icmeli C, Ozbulut O. Paradoxical therapy in conversion reaction. J Korean Med Sci 2003 Aug;18(4):581-584 [FREE Full text] [doi: 10.3346/jkms.2003.18.4.581] [Medline: 12923337]
41. Pinsof W, Breunlin D, Russell W, Lebow J, Chambers A. Integrative Systemic Therapy: Metaframeworks for Problem Solving With Individuals, Couples, and Families. New York, NY: American Psychological Association; 2018.
42. Hale D, Frusha C. MRI brief therapy: a tried and true systemic approach. J Syst Ther 2016 Jun;35(2):14-24. [doi: 10.1521/jsyt.2016.35.2.14]
43. Vitry G, Pakrosnis R, Brosseau OG, Duriez N. Effectiveness and efficiency of strategic and systemic therapy in naturalistic settings: preliminary results from a systemic practice research network (SYPRENE). J Fam Ther 2021 Apr 16;43(4):516-537. [doi: 10.1111/1467-6427.12343]
44. Smoktunowicz E, Barak A, Andersson G, Banos RM, Berger T, Botella C, et al. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. Internet Interv 2020 Sep;21:100331 [FREE Full text] [doi: 10.1016/j.invent.2020.100331] [Medline: 32577404]
45. Knopp K, Schnitzer JS, Khalifian C, Grubbs K, Morland LA, Depp C. Digital interventions for couples: state of the field and future directions. Couple Family Psychol Res Pract 2023 Dec;12(4):201-217. [doi: 10.1037/cfp0000213]
46. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
47. Helps S, Le Coyte Grinney M. Synchronous digital couple and family psychotherapy: a meta‐narrative review. J Fam Ther 2021 Jan 27;43(2):185-214. [doi: 10.1111/1467-6427.12333]
48. Baumel A, Pawar A, Kane JM, Correll CU. Digital parent training for children with disruptive behaviors: systematic review and meta-analysis of randomized trials. J Child Adolesc Psychopharmacol 2016 Oct;26(8):740-749. [doi: 10.1089/cap.2016.0048] [Medline: 27286325]
49. Leijten P, Gardner F, Melendez-Torres GJ, van Aar J, Hutchings J, Schulz S, et al. Meta-analyses: key parenting program components for disruptive child behavior. J Am Acad Child Adolesc Psychiatry 2019 Feb;58(2):180-190. [doi: 10.1016/j.jaac.2018.07.900] [Medline: 30738545]
50. Carr A. Family therapy and systemic interventions for child‐focused problems: the current evidence base. J Fam Ther 2018 Apr 27;41(2):153-213. [doi: 10.1111/1467-6427.12226]
51. Gelin Z, Cook-Darzens S, Simon Y, Hendrick S. Two models of multiple family therapy in the treatment of adolescent anorexia nervosa: a systematic review. Eat Weight Disord 2016 Mar;21(1):19-30. [doi: 10.1007/s40519-015-0207-y] [Medline: 26223191]
52. Pinquart M, Oslejsek B, Teubert D. Efficacy of systemic therapy on adults with mental disorders: a meta-analysis. Psychother Res 2016;26(2):241-257. [doi: 10.1080/10503307.2014.935830] [Medline: 25032487]
53. Köhnen M, Kriston L, Härter M, Dirmaier J, Liebherz S. Rationale and design of a systematic review: effectiveness and acceptance of technology-based psychological interventions in different clinical phases of depression management. BMJ Open 2019 Mar 27;9(3):e028042 [FREE Full text] [doi: 10.1136/bmjopen-2018-028042] [Medline: 30918040]
54. Hair K, Bahor Z, Macleod M, Liao J, Sena E. The automated systematic search deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. BMC Biol 2023 Sep 07;21(1):189 [FREE Full text] [doi: 10.1186/s12915-023-01686-z] [Medline: 37674179]
55. Higgins JP, Savović J, Page MJ, Elbers RS, Sterne JA. Assessing risk of bias in a randomized trial. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: The Cochrane Collaboration; 2021.
56. Murry VM, Berkel C, Inniss-Thompson MN, Debreaux ML. Pathways for African American success: results of three-arm randomized trial to test the effects of technology-based delivery for rural African American families. J Pediatr Psychol 2019 Apr 01;44(3):375-387 [FREE Full text] [doi: 10.1093/jpepsy/jsz001] [Medline: 30865782]
57. Riley AR, Duke DC, Freeman KA, Hood KK, Harris MA. Depressive symptoms in a trial behavioral family systems therapy for diabetes: a post hoc analysis of change. Diabetes Care 2015 Aug;38(8):1435-1440. [doi: 10.2337/dc14-2519] [Medline: 26015558]
58. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
59. Murry VM, Kettrey HH, Berkel C, Inniss-Thompson MN. The pathways for African American success: does delivery platform matter in the prevention of HIV risk vulnerability among youth? J Adolesc Health 2019 Aug;65(2):255-261. [doi: 10.1016/j.jadohealth.2019.02.013] [Medline: 31043346]
60. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
61. McGuinness LA, Higgins JP. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan 06;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]

62. Drahota A, Beller E. RevMan calculator. The Cochrane Collaboration. URL: https://training.cochrane.org/resource/revman-calculator [accessed 2024-04-29]
63. Deeks JJ, Higgins JP, Altman DG, McKenzie JE, Veroniki AA. Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. New York, NY: The Cochrane Collaboration; 2022.
64. Truttmann S, Philipp J, Zeiler M, Franta C, Wittek T, Merl E, et al. Long-term efficacy of the workshop vs. online SUCCEAT (supporting carers of children and adolescents with eating disorders) intervention for parents: a quasi-randomised feasibility trial. J Clin Med 2020 Jun 18;9(6):1912 [FREE Full text] [doi: 10.3390/jcm9061912] [Medline: 32570930]
65. Kurowski BG, Taylor HG, McNally KA, Kirkwood MW, Cassedy A, Horn PS, et al. Online family problem-solving therapy (F-PST) for executive and behavioral dysfunction after traumatic brain injury in adolescents: a randomized, multicenter, comparative effectiveness clinical trial. J Head Trauma Rehabil 2020;35(3):165-174 [FREE Full text] [doi: 10.1097/HTR.0000000000000545] [Medline: 31834062]
66. Wade SL, Cassedy AE, McNally KA, Kurowski BG, Kirkwood MW, Stancin T, et al. A randomized comparative effectiveness trial of family-problem-solving treatment for adolescent brain injury: parent outcomes from the coping with head injury through problem solving (CHIPS) study. J Head Trauma Rehabil 2019;34(6):E1-E9. [doi: 10.1097/HTR.0000000000000487] [Medline: 31033747]
67. Wade SL, Cassedy AE, Sklut M, Taylor HG, McNally KA, Kirkwood MW, et al. The relationship of adolescent and parent preferences for treatment modality with satisfaction, attrition, adherence, and efficacy: the coping with head injury through problem-solving (CHIPS) study. J Pediatr Psychol 2019 Apr 01;44(3):388-401. [doi: 10.1093/jpepsy/jsy087] [Medline: 30452665]
68. Wade SL, Cassedy AE, Taylor HG, McNally KA, Kirkwood MW, Stancin T, et al. Adolescent quality of life following family problem-solving treatment for brain injury. J Consult Clin Psychol 2019 Nov;87(11):1043-1055. [doi: 10.1037/ccp0000440] [Medline: 31556652]
69. Greenwood H, Krzyzaniak N, Peiris R, Clark J, Scott AM, Cardona M, et al. Telehealth versus face-to-face psychotherapy for less common mental health conditions: systematic review and meta-analysis of randomized controlled trials. JMIR Ment Health 2022 Mar 11;9(3):e31780 [FREE Full text] [doi: 10.2196/31780] [Medline: 35275081]
70. Page MJ, Higgins JP, Sterne JA. Assessing risk of bias due to missing results in a synthesis. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: The Cochrane Collaboration; 2022.
71. Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. Psychother Res 2018 May 09;28(3):347-355. [doi: 10.1080/10503307.2017.1411628] [Medline: 29224503]
72. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. J Educ Behave Stat 2005 Sep 01;30(3):261-293. [doi: 10.3102/10769986030003261]
73. McAloon J, de la Poer Beresford K. Online behavioral parenting interventions for disruptive behavioral disorders: a PRISMA based systematic review of clinical trials. Child Psychiatry Hum Dev 2023 Apr 24;54(2):379-396. [doi: 10.1007/s10578-021-01253-z] [Medline: 34561755]
74. Boston P. Systemic family therapy and the influence of post-modernism. Adv Psychiatr Treat 2018 Jan 02;6(6):450-457. [doi: 10.1192/apt.6.6.450]
75. Lorås L, Bertrando P, Ness O. Researching systemic therapy history: in search of a definition. J Fam Psychother 2017 Feb 17;28(2):134-149. [doi: 10.1080/08975353.2017.1285656]
76. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ 2020 Jun 04;369:m1714 [FREE Full text] [doi: 10.1136/bmj.m1714] [Medline: 32499297]
77. Hasselberg MJ. The digital revolution in behavioral health. J Am Psychiatr Nurses Assoc 2020 Oct 05;26(1):102-111. [doi: 10.1177/1078390319879750] [Medline: 31587624]
78. Greene CJ, Morland LA, Durkalski VL, Frueh BC. Noninferiority and equivalence designs: issues and implications for mental health research. J Trauma Stress 2008 Oct;21(5):433-439 [FREE Full text] [doi: 10.1002/jts.20367] [Medline: 18956449]
79. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]

Abbreviations

BFST-D: Behavioral Family Systems Therapy for Diabetes
CPP: constituent primary part
F-PST: Family Problem-Solving Therapy
MD: mean difference
PAAS: Pathways for African American Success

PRISMA: Preferred Reporting Item for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB: risk of bias
ST: systemic therapy
SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria

Edited by T de Azevedo Cardoso; submitted 07.03.23; peer-reviewed by M Rollwage, R Halkes; comments to author 01.06.23; revised version received 18.07.23; accepted 08.08.24; published 24.02.25.

Please cite as:
Erasmus P, Borrmann M, Becker J, Kuchinke L, Meinlschmidt G
Comparing Digital Versus Face-to-Face Delivery of Systemic Psychotherapy Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials
Interact J Med Res 2025;14:e46441
URL: https://www.i-jmr.org/2025/1/e46441
doi:10.2196/46441
PMID:

Review

Global Evidence on the Sustainability of Telemedicine in Outpatient and Primary Care During the First 2 Years of the COVID-19 Pandemic: Scoping Review Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) Framework

Daniela Valdes[1], MSc; Ankit Shanker[2], MSc; Ghofran Hijazi[1], MSc; Daniel Opoku Mensah[2], PhD; Tahir Bockarie[2], PhD; Ioana Lazar[3], MSc; Siti Aishah Ibrahim[2], MSc; Hamid Zolfagharinia[4], MSc; Rob Procter[1,5], PhD; Rachel Spencer[2], MD, PhD; Jeremy Dale[2], MA, MB, PhD; Armina Paule[2], MSc; Liam Jonathon Medlin[6], BSc; Keerthana Tharuvara Kallottil[3], MSc

[1]Department of Computer Science, University of Warwick, Coventry, United Kingdom

[2]Warwick Medical School, University of Warwick, Coventry, United Kingdom

[3]Warwick Manufacturing Group, University of Warwick, Coventry, United Kingdom

[4]Research & Innovation, Birmingham Community Healthcare Foundation Trust, National Health Service, Birmingham, United Kingdom

[5]The Alan Turing Institute for Data Science and AI, London, United Kingdom

[6]Department of Chemistry, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Daniela Valdes, MSc
Department of Computer Science
University of Warwick
6 Lord Bhattacharyya Way
Coventry, CV4 7AL
United Kingdom
Phone: 44 7943058485
Email: daniela.valdes@warwick.ac.uk

Abstract

Background: The rapid implementation of telemedicine during the early stages of the COVID-19 pandemic raises questions about the sustainability of this intervention at the global level.

Objective: This research examines the patient experience, health inequalities, and clinician-patient relationship in telemedicine during the COVID-19 pandemic's first 2 years, aiming to identify sustainability factors.

Methods: This study was based on a prepublished protocol using the Joanna Briggs Institute (JBI) methodology for scoping reviews. We included academic and gray literature published between March 2020 and March 2022 according to these criteria: (1) population (any group); (2) concepts (patient experience, clinician-patient relationship, health inequalities); (3) context (telemedicine in primary and outpatient care); (4) excluding studies pertaining to surgery, oncology, and (inpatient) psychiatry. We searched Ovid Medline/PubMed (January 1, 2022), Web of Science (March 19, 2022), Google/Google Scholar (February and March 2022), and others. The risk of bias was not assessed as per guidance. We used an analysis table for the studies and color-coded tabular mapping against a health care technology adoption framework to identify sustainability (using double-blind extraction).

Results: Of the 134 studies that met our criteria, 49.3% (66/134) reported no specific population group. Regarding the concepts, 41.8% (56/134) combined 2 of the concepts studied. The context analysis identified that 56.0% (75/134) of the studies referred to, according to the definition in the United Kingdom, an outpatient (ambulatory care) setting, and 34.3% (46/134) referred to primary care. The patient experience analysis reflected positive satisfaction and sustained access during lockdowns. The clinician-patient relationship impacts were nuanced, affecting interaction and encounter quality. When mapping to the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework, 81.3% (109/134) of the studies referenced the innovation's

sustainability. Although positive overall, there were some concerns about sustainability based on quality, eHealth literacy, and access to health care for vulnerable migrants and the uninsured.

Conclusions: We identified confusion between the concepts of patient experience and patient satisfaction; therefore, future research could focus on established frameworks to qualify the patient experience across the whole pathway and not just the remote encounter. As expected, our research found mainly descriptive analyses, so there is a need for more robust evidence methods identifying impacts of changes in treatment pathways. This study illustrates modern methods to decolonize academic research by using gray literature extracts in other languages. We acknowledge that the use of Google to identify gray literature at the global level and in other languages has implications on reproducibility. We did not consider synchronous text-based communication.

Trial Registration: Open Science Framework 4z5ut; https://osf.io/4z5ut/

(Interact J Med Res 2025;14:e45367) doi:10.2196/45367

KEYWORDS

pandemic; primary care; outpatients; telemedicine; ambulatory care; global health; patient experience; NASSS; clinician-patient relationship; health inequalities; gray literature; PRISMA

Introduction

Following the World Health Organization (WHO) announcement declaring COVID-19 a pandemic on March 11, 2020, the organization recommended telemedicine as one of the first critical interventions to minimize demands on stretched supplies of personal protective equipment [1,2]. Although telehealth was not new as a delivery mode, there were great expectations particularly around the use of video consultations in this context. In Africa, telemedicine held promise, as it rationalized human resources allowing national or international experts to relay advice to other clinicians [3]; in the United Kingdom, the pandemic was deemed a “burning platform” to propel the UK National Health Service (NHS) toward widespread adoption of video consultations [4]. Most medical specialties responded to the WHO recommendation with rapid changes in service delivery toward telemedicine (both telephone and a new medium—video consultations) in primary and secondary care across the globe.

In the United Kingdom, 15 months into the pandemic, the academic community and political groups raised questions around the sustainability and impacts of the move toward telemedicine, building upon the learning of the past year toward a “new normal,” particularly as social distancing and lockdown measures were removed [5]. A 2021 report by the UK Health and Social Care Committee defined telemedicine as a “welcomed and positive innovation” overall while highlighting concerns by various national organizations on its impact on health inequalities in terms of exclusion (lack of access), quality, and patient safety. The committee reported important consensus from recognized local institutions such as the Health Foundation and the Kings Fund and patient organizations such as National Voices and Healthwatch on the need for additional research to assess the future of telemedicine in a patient-centered way [6].

In previous protocols [7,8], we relayed how prepandemic evidence synthesis identified several barriers and objections that hindered telemedicine uptake, including technology, workload, and confidentiality [9], as well as concerns regarding appropriateness [10]. However, at the outset of the pandemic, these objections were rapidly overcome, supported by major regulatory and financial enablers [11,12]. Given the considerable incentives and support toward the implementation of telemedicine, there are concerns about the risks of losing some of the advantages of this mode of delivery in a postpandemic future [13], particularly once incentives are no longer in place [14,15]. These concerns apply to not only the United Kingdom, Canada, or the United States but also sub-Saharan Africa [16] and Latin America [17] where considerable barriers persist and there has been more limited evidence of uptake.

The purpose of this scoping review was to explore the global evidence (both academic and nonacademic) surrounding the rapid adoption of telemedicine in outpatient and primary care settings during the first 2 years of the COVID-19 pandemic to identify how elements related to patient experience, clinician-patient relationship, and health inequalities support (or take away from) the sustainability of this delivery model.

Methods

Protocol

This review was conducted according to an a priori published protocol [8] following the Joanna Briggs Institute (JBI) methodology for scoping reviews, updated guidance, and data extraction guidance [18-20]. This last guidance clarifies that a review can focus on the most relevant section of a document for analysis, without having to review whole studies in scoping reviews [20]. We outlined deviations from the original protocol and the methodology in Multimedia Appendix 1. A key contribution of this study hinges on the methods used to search and extract gray literature across a wider set of countries as to achieve truly global representation.

Inclusion Criteria

Population, Concept, Context Principle

The inclusion criteria used the Population, Concept, Context (PCC) principle. We classified each article against the PCC framework. Namely, for each document, we sought to identify (1) the population group (if any) to which the document referred, (2) the concept to which it referred (patient experience, clinician-patient relationship, health inequalities, or a combination of these), and (3) in which (clinical) context telemedicine was being used (outpatient, primary care, or particular specialties).

Population

The review focused on primary care services offered to the general population. Studies focusing on specific population groups or those with particular conditions within a particular country or geographical area were included.

Concept

Although the key concept under consideration was the adoption of "telemedicine," as defined in the Introduction, we narrowed our inclusion criteria to the sustainability of the interventions, focusing on patient experience, health inequalities, and clinician-patient relationship. Telemedicine has been defined as per our protocol [8] using academic literature [21] and the WHO [22], which defines it as "The delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies [...]." Although we focused on telephone and video as communication technologies, we thank an anonymous peer reviewer who identified that text messaging also falls in this category (albeit not included in our review).

Context

The context was primary care services provided during the COVID-19 pandemic in any setting or country during the first 2 years of the pandemic (March 2020-March 2022).

Types of Sources

This scoping review considered quantitative, qualitative, and mixed methods study designs for inclusion. In addition, systematic reviews, protocols, other documents, and commentaries or opinions were considered for inclusion in the proposed scoping review. These commentaries or other documents might have appeared in peer-reviewed journals or other gray literature such as industry magazines or reports [23,24].

Search Strategy

This section summarizes our prepublished protocol [8]. We structured this section by first explaining the selection and identification of search terms (both in English and other languages), then how we used those terms to search for academic and gray literature. To note, the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [25] was used to structure our review protocol [8].

Identifying Search Terms

Our search terms, which were in English, centered around telemedicine, primary care, and COVID-19 and were expanded through a limited search on Ovid Medline and Web of Science. We also sought guidance from University of Warwick librarians, who identified COVID-19 search terms from the National Institute for Health and Care Excellence (NICE) [26]. Additionally, we examined previous telemedicine protocols in general practice [27] and Primary Care Cochrane Library Protocols. For search terms in other languages, we used a 2-step process to identify relevant terms for the included non-English search results. We used Google Translate and engaged in discussions with native speakers to ensure these were accurate translations.

Search Approaches

We had 2 distinct approaches depending on whether we were searching for academic or gray literature. With the aforementioned terms, we searched the following academic databases: Ovid Medline (equivalent to PubMed [28]), Web of Science, and Google Scholar. PROSPERO and Cochrane Library were used to inform our design. PROSPERO in particular was used to check whether there was any ongoing review on the topic. All identified search terms and examples of the searches (for academic and gray literature) can be found in Multimedia Appendix 2, as per our previously published protocol [8].

We searched Ovid Medline/PubMed (January 13, 2022), Web of Science (March 19, 2022), Livivo (March 15, 2022), Scopus (March 19, 2022), PROSPERO (January 12, 2022), Cochrane Library (January 12, 2022), and Google/Google Scholar (February 2022 and March 2022).

As shown in Multimedia Appendix 2, to identify relevant gray literature published at the time of the pandemic, we used advanced Google search criteria with simplified search terms. We combined the terms telemedicine, "Primary Care," and COVID-19 with (1) "patient experience," (2) "health inequalities," or (3) "patient-clinician interaction" and asked the search engine to provide pdf-only results within the years 2020 through 2022. Although we recognize it is not fully possible to reproduce Google searches, the selection of pdf documents was aimed at identifying the most retrievable and credible gray literature [8] while, at the same time, supporting reproducibility of the analysis [23,24].

We undertook these searches in English, Chinese, Spanish, Arabic, Portuguese, Hindi, and Indonesian. To improve representation of African countries due to difficulties searching in Urdu, we undertook additional Google searches in English including country-specific results for the 5 largest African countries by population (Nigeria, Ethiopia, Democratic Republic of Congo, Egypt, and South Africa [29]). For Google searches, we selected the first 30 results by relevance, and for English-based results, we selected the top 10 results by country. We selected the first 30 results based on the literature [30-33], timeline, and budget. Further, as can be seen in the example for a failed search for Pakistan in Multimedia Appendix 2, Google searches only provided less than a handful of results when restricting by country of publication.

We followed established guidelines for analyzing non-English text [34]. We used machine translation via Google Translate to translate at least 3 paragraphs containing the key search terms (telemedicine, primary care, and any of the combinations of patient experience, clinician-patient relationship, and health inequalities). We selected 3 paragraphs that were roughly equivalent in the number of words to that of an abstract.

Screening

The authors used a single-phase, double-blind screening of abstracts and extracts based on the eligibility criteria. Screening instructions were tested by 4 authors across a sample of 50

abstracts to verify the instructions had been properly understood. The remaining articles were allocated across several combinations of pairs of authors using double-blind screening and Rayyan [35] as an aid. Any discrepancies were resolved by the pair of authors themselves and verified by the lead author before data extraction. Given the extended scope of our review, and in agreement with the prepublished protocol and the latest JBI guidance [20], there was no full-text screening as the abstracts and extracts were our main data source. Abstracts and extracts have been made available in Multimedia Appendix 3.

Data Extraction

The authors used double-blind extraction of data using an Excel table template as outlined in Multimedia Appendix 4. Following the latest JBI guidance for scoping reviews [20], we chose to focus data extraction on abstracts and extracts only, reflecting the wide research design that allowed us to accommodate the breadth of the review in terms of (1) themes; (2) a worldwide, multilanguage approach; and (3) sources. Using a data extraction tool shared in Multimedia Appendix 4, we mapped the text in tabular form against the NASSS framework [36] domains, noting some of the document sections might touch upon 1 or more domains.

Data Analysis and Presentation

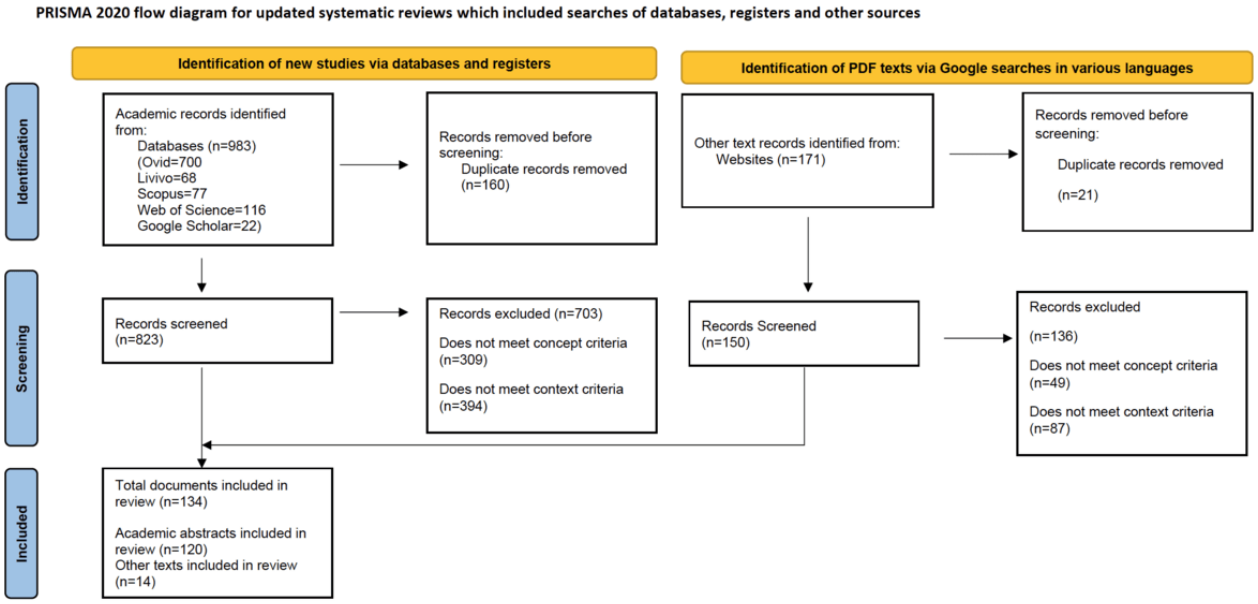
In agreement with the latest JBI methodological guidance [18-20], no critical appraisal was undertaken, and the final presentation of results consisted of the following. In the first section, the results of the search were presented in a PRISMA-ScR flow diagram [25], including a flowchart and a checklist, and we conducted a table analysis of more detailed characteristics of the included documents (Table S1 in Multimedia Appendix 5). We captured the number of articles that included certain data against the PCC framework. Understandably, in the “population” label, the numbers do not add up to the total as some categories are not mutually exclusive. A color-coded (heat map) mapping was created in tabular form against the NASSS framework’s [36] domains: (1) the condition, (2) the technology, (3) the value proposition, (4) adopters, (5) organizations, (6) wider system, and (7) embedding and adaptation over time. The heat map shows graphically the maximum and minimum numbers of references for each domain (using the average counts for double-blind data extraction and mapping). We performed a narrative analysis (including sentiment analysis) of references to the sustainability of video consultations outside of social distancing restrictions brought about by the pandemic. The results are discussed from the point of view of sociotechnical grounded theory, providing strengths and limitations of the sources and the review method itself. We included a statement of positionality in our conclusions, as well as recommendations for research and practice. We believe that reflexivity through researcher positionality is fundamental to decolonizing global health research that seeks to include voices and perspectives usually marginalized from the academic discourse [37-39].

Results

PRISMA Study Inclusion

Academic database searches identified 983 records, and 171 additional records were identified through Google searches for various languages and countries. For Google Arabic and Google Pakistan, there were no results that matched the search criteria. Of the total 1154 articles, 181 were identified as duplicates and were removed: 160 duplicates were academic articles, and 21 duplicates were identified among the nonacademic articles. After the removal of duplicates, there were 973 articles to be screened. Of these 973 articles, 823 articles were from academic databases, and 150 were from nonacademic databases. After the abstract screening and before data extraction, 703 academic and 136 nonacademic articles were excluded, leaving a total of 134 documents [29,40-171] for data extraction including 120 academic articles and 14 extracts from gray literature searches. See the PRISMA-ScR [25] flowchart (Figure 1).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart outlining the process of identification, screening, and final inclusion across various types of data sources.



Summary of Document Types

We separated the academic studies (which we identified as those having a clearly labelled abstract with introduction, aims, methods, results, and discussion sections) from the other documents (or gray literature) and found that these other documents represented 27.6% (37/134) of the selected documents. Searches in academic databases and journals provided a small group of gray literature in the form of commentaries or guidelines. Conversely, general internet searches also identified a minority of academic documents.

The document pool achieved global representation, with documents from all continents, including South America [46,110,166] and Africa [116,121]. There was, however, overrepresentation from North America (58/134, 43%), with a large proportion of documents from the United States.

In terms of the methodology in the selected documents, over 72% (96/134) of the documents used surveys, questionnaires, or interviews. The use of surveys and questionnaires was closely related to the type of design observed, with most studies being cross-sectional (96/134, 72%). Finally, regarding the telemedicine medium, the documents did not specify the medium by generally referring to “telemedicine” in 57% (77/134) of documents. More details are provided in Table S2 in Multimedia Appendix 5.

Quantitative Classification Against the PCC Framework

Table S3 in Multimedia Appendix 5 includes the characteristics of the 134 documents against the protocol’s PCC. We follow with a short commentary highlighting any documents that exemplify these findings. Table 1 shows how the documents related to the various key concepts explored (clinician-patient relationship, health inequalities, and patient experience), and the following paragraphs summarize the key findings. In addition, Figure 2 shows the results of the bibliographic keyword analysis.

In terms of population groups, a large subset of abstracts and extracts (66/134, 49.3%) reported no specific demographics nor patient characteristics (see, for example, the extract from the board report from the East Kent Hospitals NHS Foundation Trust [64] or Karacabeyli et al [95]). The main demographic characteristics reported in abstracts and extracts were age (39/134, 29.1%) and sex (20/134, 14.9%), with studies also considering sociodemographic factors. It is important to note that a particular document might have tracked more than one characteristic; see, for example, Manski-Nankervis et al [113], which tracked education status, gender, age, and whether patients spoke English at home.

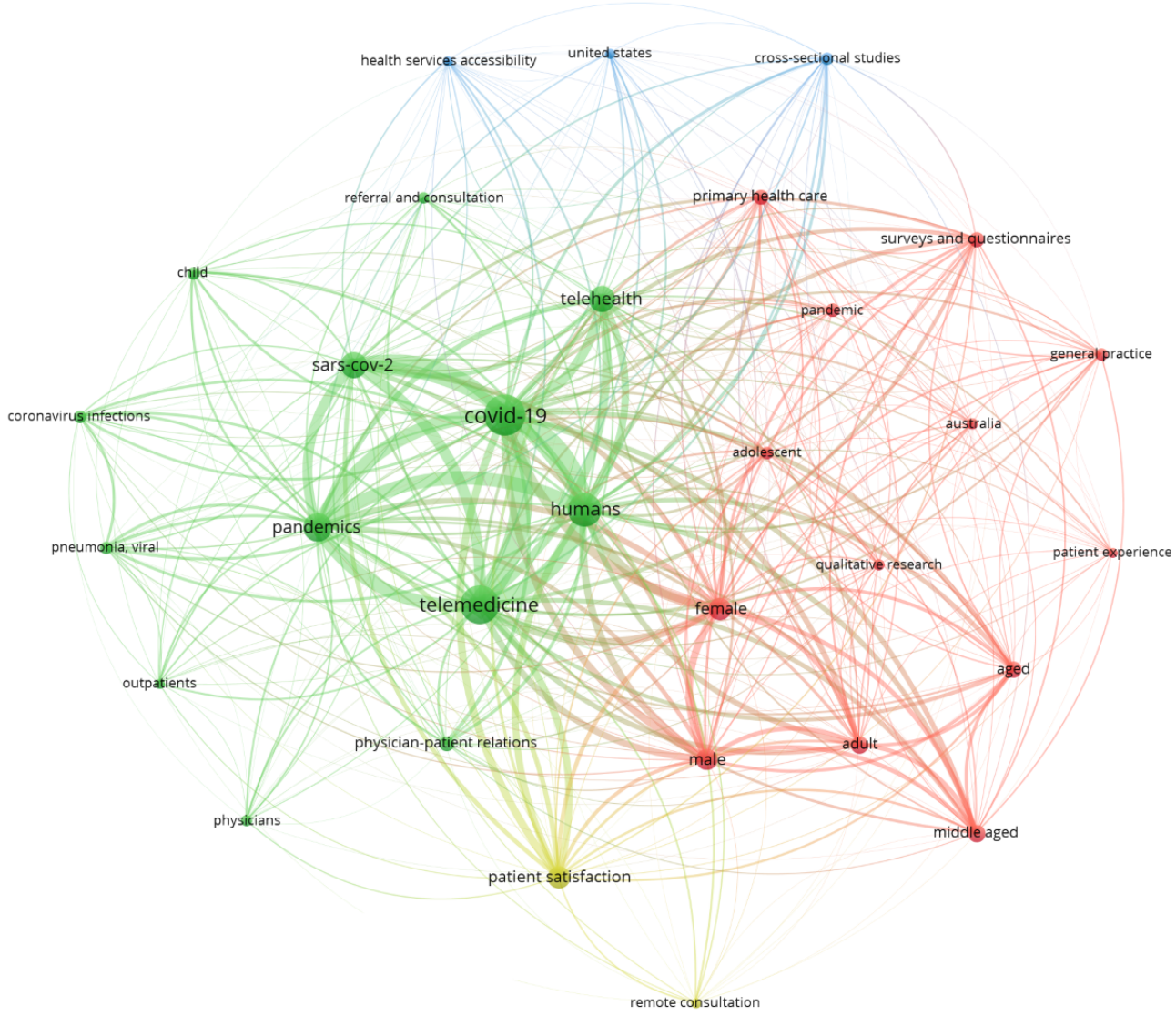
In terms of the concepts studied**,** the most popular concept was the clinician-patient relationship, reported in 28 abstracts. The majority of documents (69/134, 51.5%) combined 2 or 3 concepts. In the following paragraphs, we provide a brief summary of results across the various concepts.

Regarding patient experience, we found 26 studies solely focused on this concept. Patient experience was mostly equated with patient satisfaction and access. There were positive levels of satisfaction overall [99,121,149] and sustained access at the time of lockdowns [48,53].

Table 1. Classification of the documents against the key concepts.

Concept	Documents, n
1 concept	
Clinician-patient relationship	28
Health inequalities	11
Patient experience	26
2 concepts	
Clinician-patient relationship, health inequalities	11
Clinician-patient relationship, patient experience	29
Health inequalities, patient experience	16
All 3 concepts	
Clinician-patient relationship, patient experience, health inequalities	13

Figure 2. Bibliographic keyword analysis.



Regarding health inequalities, there were only 11 studies solely focused on this concept. We found gaps in evidence that made it difficult to pinpoint the impacts on health inequalities of specific groups, with some evidence of health inequality exacerbation for those considered to be at the fringe of the economic and health system (such as vulnerable migrants, families with high levels of economic deprivation or health care insurance exclusion, or rural communities with limited access to technology) [74,116,140].

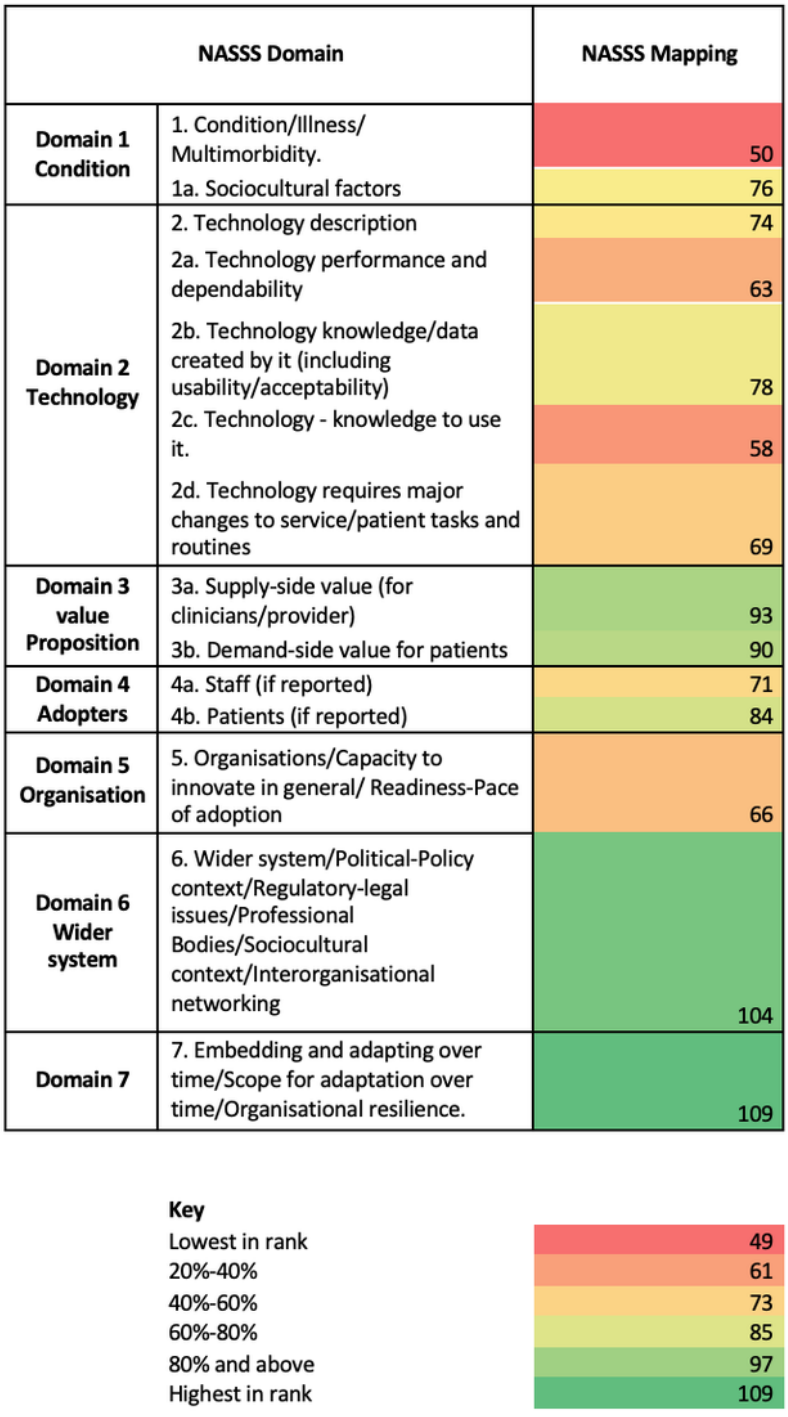
Regarding the clinician-patient relationship, there were 28 documents solely focused on this concept, mainly on the clinician-patient interaction and quality of the encounter. There was no consensus regarding whether the impact of telemedicine on the interaction had been positive, neutral, or negative [43,56,70], while there was consensus on the added value of the quality of the remote encounter for triage, follow-up consultations, or chronic condition management [128,133,155].

Finally, concerning the context, we identified 46 documents focused solely on primary care, general outpatient care featured in 19 documents, and various outpatient specialties were featured in 56 extracts.

Mapping Against the NASSS Framework

As shown in Figure 3, there was considerable variability in evidence across the NASSS [36] domains and subdomains. Domains 6 (wider system), 7 (time domain), and 3 (value proposition) had the most information. For domain 6, documents referenced the pandemic or pandemic-related lockdowns and infection control measures coupled with regulatory enablers or recommendations (such as mandatory online triage in primary care, parity in payments between face-to-face and telemedicine appointments). There was lower density of information about specific conditions that were being managed with telemedicine (subdomain 1a), knowledge needed to use telemedicine (subdomain 2c), and organizational or implementation aspects (domain 5).

Figure 3. Heat map against each of the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) domains.



Narrative Analysis of Sustainability

Roughly over one-half of the documents (75/134, 56%) identified challenges in terms of sustainability or made recommendations on how to address them. Among these, sentiment was mixed or neutral in 40 documents and positive in 31 documents, with only a small subgroup [68,89,141,148] viewing such challenges negatively as barriers to further planning and progress. Challenges and areas of consideration included general planning such as workforce training, digital resources, patient experience, and ethical issues [92] or a more focused look at the technology itself such as more effective digital platforms and increased use of home medical equipment [43]. Other issues such as patient selection were also noted with consideration to disease progression, language and cognitive ability, health literacy, and technology access [40]. The systematic review by Mbunge et al [116] of the digital technologies deployed in South Africa during the pandemic

identified a "digital divide" barrier in rural areas and advocated for better networks.

Almost one-third (38/134, 28%) of the documents set out plans for sustainable growth and further embedding, with generally positive sentiment. The report by the East Kent Hospitals NHS Foundation Trust in 2021 [64] mentioned an enhanced engagement plan to meet ambitious targets set by health care authorities for the delivery of telemedicine appointments. Tulupova et al [162] referred to plans for the creation of telemedicine guidelines and an educational program on communications in health care using digital technologies for patients to improve digital health literacy. The remaining documents either did not address the area of sustainability [42,44,49,66,95,101,106,110,111,122,132,143,153,167] or had a generally negative view on the sustainability of the intervention [63,74,80,112,169]. The negative commentary was based on concerns about quality (ie, treating musculoskeletal conditions), eHealth literacy, and access to health care for those at the fringe of the health care system coverage (such as vulnerable migrants and the uninsured).

Discussion

This section includes 4 main areas: (1) a summary of our findings, (2) a sociotechnical grounded theory research interpretation of the findings (based on Hoda [172]), (3) positioning against the wider and recent literature, and (4) strengths and limitations of our study. The section finalizes with a short conclusion.

Summary of Findings

Concerning the protocol's PCC, 49% (66/134) of the documents reported no specific population group targeted by the study (population); according to the UK definition, 55% (74/134) of the studies referred to an outpatient (ambulatory care) setting, and 34% (46/134) referred to primary care (context); and 49% (66/134) of the studies referred to only 1 of the concepts studied (concept).

Mapping to the NASSS framework [36] identified that 93% (125/134) of the studies referenced the sustainability of telemedicine with moderately positive comments.

Global Representation

We found 134 studies meeting our criteria and achieved global representation. We highlight some of the global results from limited-resource countries here.

Despite the limited evidence for African countries (2 studies [116,121]), our findings were aligned with other African reviews around insufficient evidence of use due to considerable barriers in this region [16] and a lack of "meaningful investment" in this area [173]. We concurred with Nittari et al [174] that several barriers are still present that risk the sustainability of this delivery mode beyond the pandemic.

Studies in South America [46,110,166] provide examples of effective use in outpatient settings (specifically speech and language therapy [46]), as well as reflections on how this new delivery mode in the context of the wider pandemic might have generated an element of mistrust and fear in the clinician-patient relationship [166].

Results from Middle Eastern and Asian countries [41,97,104,105] provide perspectives on the use of telemedicine in orthopedics and hematology while indicating the equivalence of audio and video consultations, an important point to inform telemedicine programs in low-income countries as a way of increasing access to health services.

Positioning Against the Wider and Recent Literature

This review can be positioned in the emerging literature of reviews around telemedicine during the pandemic period [175-183].

A key finding of our study is how patient experience was generally equated with patient satisfaction. Other studies have found these are often used interchangeably [184-186]. However, accepted definitions of patient experience go beyond satisfaction and "focus on individualized care and tailoring of services to meet patients' needs" [187]. This is related to another finding, as the studies are mainly observational with no reference to patient experience frameworks, let alone to some emerging frameworks specific to the digital patient experience [173,188]. Our findings on patient experience are aligned with another recently published review focused on the COVID-19 pandemic, with similar categories and findings [178]. In terms of the clinician-patient relationship, our findings were mixed, but recent reviews found that the relationship was "troubled" telemedicine given both patient and clinician reluctance to use [177]. However, others found the tool useful [106].

de Oliveira Andrade et al [175] focused on the role of legislation in the widespread utilization of telemedicine during the pandemic, finding that regulatory frameworks enabled telemedicine spread in areas related to ethics, reimbursement, data safety, and pandemic-related regulatory relaxation (in the United States, for instance, relaxation of interstate practice was particularly relevant [183]). We found a lack of consensus in terms of sustainability as these financial and regulatory incentives dissipated; however, supportive regulation would be a defining factor in its sustainability. Our evidence seems more nuanced than other recent reviews [177,189]. We are aware specialties like family medicine and general practice seem to have a preference for a particular medium, such as telephone instead of video [190], with more work being undertaken regarding the impacts on quality [191] or equivalence between face-to-face and remote consultations [183].

Regarding the impact on health inequalities, another review identified emerging literature on the opinions of vulnerable populations regarding telemedicine [180]. We saw references [192-195] to an emerging framework for digital health equity [196], which we expect will help address the gaps we identified in the design and evaluation of inclusive digital health care services to address the "self-reinforcing effect of digital and social exclusion" [197] and its impact on health care inequalities in access and experience. Our concerns about implications on health inequality resonate with other similar reviews for telemedicine [179].

Strengths and Limitations

Strengths

We mapped the emerging literature (gray and academic) on telemedicine during the pandemic to a well-established framework using 3 lenses (patient experience, health inequalities, and clinician-patient relationship). We identified this as the only systematic review of its kind. Only 1 other review mapped academic literature on video consultations to the NASSS framework [36] but outside the context of the pandemic [37]. Notably, we effectively reflected the experiences of non–English-speaking countries with literature across the 5 continents, so it is truly global health–oriented, with the added values and perspectives of a diverse, multidisciplinary research team.

When identifying the limited literature at the intersection of health inequalities and telemedicine, we provided a brief taxonomy of potential groups impacted differently. Recognizing the multidimensionality and intersectionality of social exclusion, we show that demographic characteristics such as age, sex, and socioeconomic factors have received some attention, but there is still very limited information and not enough to draw solid conclusions on impacts.

From a methodological standpoint, we provided additional insight on how to integrate effectively other documents and nontraditional voices and experiences into academic research with a reproducible approach.

Limitations of the Sources

The authors sought gray literature directly via Google to ensure a unified source and methodology to identify and capture experiences from non–English-speaking countries. This is a method with limited reproducibility. Researchers interested in gray literature information in English can consult the UK National Grey Literature Collection Funded by Health Education England. We are aware of additional databases with non-English literature that could be used [198]. Emerging literature has covered mostly single-center survey studies with limited sample sizes, reflective of the immediate experiences arising in the context of the pandemic.

Limitations of the Review

We have limited the depth of analysis to accommodate for the extensive scope, in accordance with the JBI guidelines for data extraction [199]. The use of document extracts using the "surrounding keyword" approach needs further development and testing, as we recognize slightly longer extracts are better at conveying enough information to support screening and analysis. We did not consider synchronous text-based communication. The methodology of scoping reviews is still skewed toward evidence from academic publications, which are biased toward researchers from North America and Europe. Although not specified in the current guidance, capping the number of results from traditional academic databases provides a more balanced representation and could reduce duplication while having limited effects on how comprehensive the findings are.

In conclusion, our discussion section has highlighted considerable variation in the emerging literature during the pandemic regarding changes in pathways toward telemedicine. We highlighted the different focus of studies focused on health inequality or outpatient care and the global representation of the studies included (which is a key strength). Of note is the finding equating patient experience with satisfaction, which reflects a potentially limited understanding of sociotechnical views of human-computer interaction and human factors in traditional health service research.

Conclusions

This section is divided into 3 parts comprising a reflection of our positionality as researchers in analyzing these themes as well as further recommendations for research and practice.

Statement of Positionality

Following the method by Pant et al [37], we frame some of our conclusions in the context of our positionality and our aims. We are a diverse group of researchers (with roots in Latin America, the Middle East, Asia, and Africa and with supervisors from the United Kingdom), but our gaze is colonially influenced by our education and current positionality in this country.

By opening our search criteria to non-English documents and gray literature, we succeeded in capturing immediate, emerging experiences at the global level, with 27.6% (37/134) of our documents having no structured abstract and classified as gray literature, with 5 documents from South America and Africa and 13 documents from Asia. The balance, however, is still very skewed toward the United States and English-speaking countries.

Our positionality and knowledge of the UK health system and legislation meant it was difficult to translate these categories to other systems, and we had to modify our parameters and analysis. For instance, although our choice of the UK diversity legislation ("protected characteristics") as a framework for categorizing health inequalities was helpful, the UK definition of "primary care" was not helpful due to its contrast with the definition set out by the WHO (for example in the work by Dimer et al [46] in Brazil, classifying speech and language therapy services as a "primary health care service," which are not available in the United Kingdom in this setting).

Recommendations for Research

From a methodological standpoint, we urge researchers looking to decolonize academic research to test and evaluate further our approach to using gray literature extracts in other languages, particularly in scoping reviews, as it provides that additional level of immediacy with the phenomenon of study. We noticed confusion on the use of patient satisfaction and patient experience, so research should focus on more robust frameworks reviewing the overall patient pathway (away from the evaluation of a remote encounter [184-186]). As expected, our research found mainly descriptive documents, so future research should focus on robust evaluation of clinical outcomes arising from changes in diagnostic and treatment pathways away from face-to-face settings. We recommend future research with a narrower approach to specific population groups and more

focused on access and outcomes (Table S1 in Multimedia Appendix 5). From a sociotechnical research perspective, we recommend future research using modern techniques (natural language processing, data mining, and sentiment analysis) focusing on categories with closer links to human-computer interaction. Following publication, we will publish our data on Open Science Framework [200].

Recommendations for Practice

The impact of telemedicine on patient safety is critical to determine the sustainability of this intervention when contrasting it with the under- or overutilization of resources [201]. If not ruling out the continued use of telemedicine, the authors outlined the importance of further research and refinements to the intervention itself. We found that models such as the Dynamic Sustainability Framework [75] might be useful to support learning and adaptation with care toward the potential disenfranchisement of some patient groups.

Acknowledgments

ChatGPT and Grammarly were used to edit the final draft. Many thanks to Cee MacDonald at Paperaid for the assistance with extract retrieval and text formatting.

This research did not receive any external funding and was undertaken as part of DV's self-funded PhD studies.

Authors' Contributions

These contributions use Contributor Roles Taxonomy (CRediT) guidelines. As this article is also a pedagogic exercise related to various individuals' studies, it was agreed that all voluntary contributors would be set as co-authors regardless of the degree of contribution.

This article contributes to DV's PhD studies at the University of Warwick, and DV contributed to the study conceptualization, data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), project administration (lead), software, validation, visualization, writing of the original manuscript draft (lead), and manuscript review and editing. AS contributed to the data curation (lead), formal analysis (support), investigation (support), methodology (equal), project administration (support), software (support), visualization (support), writing of the original manuscript draft (support), and copyediting (lead). DOM and IL contributed to the data curation (support), investigation (support), and methodology (equal). GH and HZ contributed to the formal analysis (support), visualization (support), and writing of the original manuscript draft (support). TB contributed to the data curation (support), investigation (support), methodology (equal), and manuscript review and editing (support). SAI contributed to the data curation (support), formal analysis (support), and methodology (support). RP contributed to study supervision (lead), validation (lead), and manuscript review and editing (support). RS contributed to study supervision (support), manuscript review and editing (lead), and validation (lead). JD contributed to study supervision (support). AP contributed to validation of data extraction (support) and preparation of the final draft. LJM and KTK contributed to the data curation (support).

Conflicts of Interest

At the time of publication, DV, TB, and HZ worked in the UK National Health Service in the United Kingdom. The research was undertaken independently of DV's, TB's, and HZ's occupations and does not represent the views of their employers. The remaining authors declare no conflict of interest.

Multimedia Appendix 1
Deviations from the protocol.
[DOCX File , 17 KB - ijmr_v14i1e45367_app1.docx]

Multimedia Appendix 2
Example searches.
[DOCX File , 26 KB - ijmr_v14i1e45367_app2.docx]

Multimedia Appendix 3
Abstracts-extracts used in the study.
[XLSX File (Microsoft Excel File), 270 KB - ijmr_v14i1e45367_app3.xlsx]

Multimedia Appendix 4
Data collection template.
[XLSX File (Microsoft Excel File), 460 KB - ijmr_v14i1e45367_app4.xlsx]

Multimedia Appendix 5
Summary characteristics of the included studies.
[DOCX File , 66 KB - ijmr_v14i1e45367_app5.docx]

Multimedia Appendix 6
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) Checklist.
[PDF File (Adobe PDF File), 101 KB - ijmr_v14i1e45367_app6.pdf]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://tinyurl.com/tn4rtbse [accessed 2021-02-20]
2. Rational use of personal protective equipment (PPE) for coronavirus disease (COVID-19). World Health Organization. 2020 Mar 19. URL: https://apps.who.int/iris/bitstream/handle/10665/331498/WHO-2019-nCoV-IPCPPE_use-2020.2-eng.pdf [accessed 2021-06-08]
3. Chersich MF, Gray G, Fairlie L, Eichbaum Q, Mayhew S, Allwood B, et al. COVID-19 in Africa: care and protection for frontline healthcare workers. Global Health 2020 May 15;16(1):46 [FREE Full text] [doi: 10.1186/s12992-020-00574-3] [Medline: 32414379]
4. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
5. Wherton J, Greenhalgh T. Evaluation of the Near Me video consulting service in Scotland during COVID-19, 2020: Summary Report. Scottish Government. 2021 Mar. URL: https://tinyurl.com/3nmavy8m [accessed 2024-11-06]
6. The NHS: lessons learnt and building for the future. UK Parliament. URL: https://publications.parliament.uk/pa/cm5801/cmselect/cmhealth/320/32009.htm [accessed 2021-06-12]
7. Valdes D, Alqazlan L, Procter R, Dale J. Global evidence on the rapid adoption of telemedicine in primary care during the first year of the COVID-19 pandemic: a scoping review protocol. Research Square. 2021 Mar 16. URL: https://assets-eu.researchsquare.com/files/rs-302879/v1/5668b4f9-5d65-4353-93db-32249840ca3d.pdf [accessed 2024-11-27]
8. Valdes D, Alqazlan L, Procter R, Dale J. Global evidence on the rapid adoption of telemedicine in primary care during the first 2 years of the COVID-19 pandemic: a scoping review protocol. Syst Rev 2022 Jun 19;11(1):124 [FREE Full text] [doi: 10.1186/s13643-022-01934-3] [Medline: 35718770]
9. Rodgers M, Raine G, Thomas S, Harden M, Eastwood A. Informing NHS policy in ‘digital-first primary care’: a rapid evidence synthesis. Health Serv Deliv Res 2019 Dec;7(41):1-124. [doi: 10.3310/hsdr07410] [Medline: 31869020]
10. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Mar 17;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20x101020]
11. Guidance and standard operating procedures: general practice in the context of coronavirus (COVID-19) Version 3.4. Sheffield Local Medical Committee. 2020 Aug 10. URL: https://tinyurl.com/2s362d9z [accessed 2024-11-06]
12. Gausvik C, Jabbarpour Y. COVID-19 Timeline: Centers for Medicare and Medicaid Services (CMS) changes and primary care support were not enough to prevent practice losses. J Am Board Fam Med 2021 Feb;34(Suppl):S7-S9 [FREE Full text] [doi: 10.3122/jabfm.2021.S1.200305] [Medline: 33622808]
13. Southgate G, Yassaee AA, Harmer MJ, Livesey H, Pryde K, Roland D. Use of telemedicine in pediatric services for 4 representative clinical conditions: scoping review. J Med Internet Res 2022 Oct 26;24(10):e38267 [FREE Full text] [doi: 10.2196/38267] [Medline: 36287609]
14. Weiner S. What happens to telemedicine after COVID-19? Association of American Medical Colleges. 2021 Oct 21. URL: https://www.aamc.org/news-insights/what-happens-telemedicine-after-covid-19 [accessed 2022-11-26]
15. Doarn CR. The receding tide: is telemedicine in retreat? Telemed J E Health 2022 Nov 01;28(11):1553-1554. [doi: 10.1089/tmj.2022.29082.editorial] [Medline: 36538750]
16. Dodoo JE, Al-Samarraie H, Alzahrani AI. Telemedicine use in Sub-Saharan Africa: Barriers and policy recommendations for Covid-19 and beyond. Int J Med Inform 2021 Jul;151:104467 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104467] [Medline: 33915421]
17. Pierce W, Schroeder D, Suchecki R. Telehealth in Latin America: progress, challenges, and opportunities in the face of COVID-19. Telehealth Med Today 2021 Jan 28;6(1):1 [FREE Full text] [doi: 10.30953/tmt.v6.238]
18. Aromataris E, Munn Z. JBI manual for evidence synthesis. JBI 2020:1 [FREE Full text] [doi: 10.46658/JBIMES-20-01]
19. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
20. Pollock D, Peters M, Khalil H, McInerney P, Alexander L, Tricco A, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 01;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]

21. Olszewski C, Thomson S, Strauss L, Graham R, Ezzeddine M, Dodenhoff K, et al. Patient experiences with ambulatory telehealth in neurology: results of a mixed-methods study. Neurol Clin Pract 2021 Dec;11(6):484-496 [FREE Full text] [doi: 10.1212/CPJ.0000000000001072] [Medline: 34992956]
22. WHO Global Observatory for eHealth. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth. World Health Organization. 2010. URL: https://iris.who.int/handle/10665/44497 [accessed 2024-11-06]
23. Adams RJ, Smart P, Huff AS. Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies. Int J Management Reviews 2016 Apr 19;19(4):432-454. [doi: 10.1111/ijmr.12102]
24. Tyndall J. How low can you go? Toward a hierarchy of Grey Literature. 2008 Presented at: Dreaming 08 – Australian Library and Information Association Biennial Conference; September 2-5, 2008; Alice Springs, NT, Australia URL: https://core.ac.uk/download/pdf/14931482.pdf [doi: 10.14325/mississippi/9781496828118.003.0029]
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Interim process and methods for developing rapid guidelines on COVID-19. NICE. 2020 Mar 20. URL: https://www.nice.org.uk/process/pmg35 [accessed 2024-11-06]
27. Downes MJ, Mervin MC, Byrnes JM, Scuffham PA. Telemedicine for general practice: a systematic review protocol. Syst Rev 2015 Oct 05;4(1):134 [FREE Full text] [doi: 10.1186/s13643-015-0115-2] [Medline: 26597992]
28. How to compare Ovid MEDLINE and PubMed. Wolters Kluwer Health. 2019 Nov. URL: https://tinyurl.com/442zne52 [accessed 2024-11-06]
29. Assenza G, Lanzone J, Brigo F, Coppola A, Di Gennaro G, Di Lazzaro V, et al. Epilepsy care in the time of COVID-19 pandemic in Italy: risk factors for seizure worsening. Front Neurol 2020;11:737 [FREE Full text] [doi: 10.3389/fneur.2020.00737] [Medline: 32719655]
30. Garousi V, Felderer M, Mäntylä MV. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. Information and Software Technology 2019 Feb;106:101-121. [doi: 10.1016/j.infsof.2018.09.006]
31. Altingovde IS, Demir E, Can F, Ulusoy Ö. Site-based dynamic pruning for query processing in search engines. 2008 Presented at: SIGIR '08: 31st annual international ACM SIGIR conference on Research and development in information retrieval; July 20-24, 2008; Singapore, Singapore. [doi: 10.1145/1390334.1390543]
32. Szmuda T, Özdemir C, Ali S, Singh A, Syed M, Słoniewski P. Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study. Public Health 2020 Aug;185:21-25 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.041] [Medline: 32516624]
33. Beel J, Gipp B. Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). 2009 Presented at: Sixth International Conference on Information Technology: New Generations; April 27-29, 2009; Las Vegas, NV. [doi: 10.1109/itng.2009.317]
34. Petticrew M, Roberts H. Systematic Reviews in the Social Sciences: A Practical Guide. Hoboken, NJ: John Wiley & Sons; 2008.
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
36. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
37. Pant I, Khosla S, Lama JT, Shanker V, AlKhaldi M, El-Basuoni A, et al. Decolonising global health evaluation: Synthesis from a scoping review. PLOS Glob Public Health 2022 Nov 16;2(11):e0000306 [FREE Full text] [doi: 10.1371/journal.pgph.0000306] [Medline: 36962490]
38. Kumar M, Atwoli L, Burgess RA, Gaddour N, Huang KY, Kola L, et al. What should equity in global health research look like? The Lancet 2022 Jul;400(10347):145-147. [doi: 10.1016/s0140-6736(22)00888-1]
39. van Wyk A. Decolonising research methods and knowledge: implications for stepping into a post-doctoral role. Health & Social Care Workforce. 2022 Aug 02. URL: https://tinyurl.com/34xcnf7c [accessed 2024-11-06]
40. Adams L, Lester S, Hoon E, van der Haak H, Proudman C, Hall C, et al. Patient satisfaction and acceptability with telehealth at specialist medical outpatient clinics during the COVID-19 pandemic in Australia. Intern Med J 2021 Jul;51(7):1028-1037 [FREE Full text] [doi: 10.1111/imj.15205] [Medline: 34213046]
41. Alhajri N, Simsekler MCE, Alfalasi B, Alhashmi M, AlGhatrif M, Balalaa N, et al. Physicians' attitudes toward telemedicine consultations during the COVID-19 pandemic: cross-sectional study. JMIR Med Inform 2021 Jun 01;9(6):e29251 [FREE Full text] [doi: 10.2196/29251] [Medline: 34001497]
42. Alharbi KG, Aldosari MN, Alhassan AM, Alshallal KA, Altamimi AM, Altulaihi BA. Patient satisfaction with virtual clinic during Coronavirus disease (COVID-19) pandemic in primary healthcare, Riyadh, Saudi Arabia. J Family Community Med 2021;28(1):48-54 [FREE Full text] [doi: 10.4103/jfcm.JFCM_353_20] [Medline: 33679189]

43. Aliberti GM, Bhatia R, Desrochers LB, Gilliam EA, Schonberg MA. Perspectives of primary care clinicians in Massachusetts on use of telemedicine with adults aged 65 and older during the COVID-19 pandemic. Prev Med Rep 2022 Apr;26:101729 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101729] [Medline: 35155085]
44. Al-Iede M, Waters K, Aleidi S, Alqutawneh B, Alnawaiseh H, Alshraideh A, et al. Impact of COVID-19 lockdown on children with asthma in Jordan: a parental questionnaire. BMJ Paediatr Open 2021;5(1):e001136 [FREE Full text] [doi: 10.1136/bmjpo-2021-001136] [Medline: 34222681]
45. Altulaihi BA, Alharbi KG, Alhassan AM, Altamimi AM, Al Akeel MA. Physician's perception toward using telemedicine during COVID-19 pandemic in King Abdulaziz Medical City, Riyadh, Saudi Arabia. Cureus 2021 Jul;13(7):e16107 [FREE Full text] [doi: 10.7759/cureus.16107] [Medline: 34350074]
46. Dimer NA, do Canto-Soares N, Dos Santos-Teixeira L, de Goulart BNG. The COVID-19 pandemic and the implementation of telehealth in speech-language and hearing therapy for patients at home: an experience report. Codas 2020;32(3):e20200144 [FREE Full text] [doi: 10.1590/2317-1782/20192020144] [Medline: 32578694]
47. Ball E, Rivas C, Khan R. If virtual gynecology clinics are here to stay, we need to include everyone. AJOG Glob Rep 2022 Feb;2(1):100043 [FREE Full text] [doi: 10.1016/j.xagr.2021.100043] [Medline: 34909705]
48. Barkai G, Gadot M, Amir H, Menashe M, Shvimer-Rothschild L, Zimlichman E. Patient and clinician experience with a rapidly implemented large-scale video consultation program during COVID-19. Int J Qual Health Care 2021 Feb 20;33(1):1 [FREE Full text] [doi: 10.1093/intqhc/mzaa165] [Medline: 33313891]
49. Belenkiy IG. COVID-19 challenge: what has been done and what must be done? Traumatology and Orthopedics of Russia 2020 Jul 06;26(2):15-19. [doi: 10.21823/2311-2905-2020-26-2-15-19]
50. Berg WT, Goldstein M, Melnick AP, Rosenwaks Z. Clinical implications of telemedicine for providers and patients. Fertil Steril 2020 Dec;114(6):1129-1134 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.10.048] [Medline: 33280717]
51. Bergman D, Bethell C, Gombojav N, Hassink S, Stange KC. Physical distancing with social connectedness. Ann Fam Med 2020 May 01;18(3):272-277 [FREE Full text] [doi: 10.1370/afm.2538] [Medline: 32393566]
52. Byrne E, Watkinson S. Patient and clinician satisfaction with video consultations during the COVID-19 pandemic: an opportunity for a new way of working. J Orthod 2021 Mar;48(1):64-73. [doi: 10.1177/1465312520973677] [Medline: 33251951]
53. Chesnel C, Hentzen C, Le Breton F, Turmel N, Tan E, Haddad R, et al. Efficiency and satisfaction with telephone consultation of follow-up patients in neuro-urology: Experience of the COVID-19 pandemic. Neurourol Urodyn 2021 Mar;40(3):929-937 [FREE Full text] [doi: 10.1002/nau.24651] [Medline: 33675263]
54. Chiang C, Halker Singh R, Lalvani N, Shubin Stein K, Henscheid Lorenz D, Lay C, et al. Patient experience of telemedicine for headache care during the COVID-19 pandemic: An American Migraine Foundation survey study. Headache 2021 May;61(5):734-739 [FREE Full text] [doi: 10.1111/head.14110] [Medline: 34021595]
55. La Telemedicina Cara a Cara. Comité de ética asistencial de Atención Primaria de Bizkaia. 2020 Dec. URL: https://osalde.org/wp-content/uploads/2021/01/Telemedicina-aspectos-%C3%A9ticos-CEAAPB.pdf [accessed 2022-10-26]
56. Conway J, Krieger P, Hasanaj L, Sun L, Scharf JM, Odel JG, et al. Telemedicine evaluations in neuro-ophthalmology during the COVID-19 pandemic: patient and physician surveys. J Neuroophthalmol 2021 Sep 01;41(3):356-361. [doi: 10.1097/WNO.0000000000001370] [Medline: 34415269]
57. Curtis M, Duncan R, Jing M, Kim A, Lu VT, Redshaw J, et al. "Not a perfect situation, but..." A single-practice survey of patient experience of phone consultations during COVID-19 Alert Level 4 in New Zealand. N Z Med J 2021 Oct 22;134(1544):35-48. [Medline: 34695091]
58. Day K, Hunter I, Rao V, Russell G, Roskvist R, Moir F, et al. Survey protocol for exploring video and phone use in Aotearoa New Zealand general practice: considerations for future telehealth. BMJ Health Care Inform 2021 Mar;28(1):1 [FREE Full text] [doi: 10.1136/bmjhci-2020-100309] [Medline: 33753326]
59. Dempsey CM, Serino-Cipoletta JM, Marinaccio BD, O'Malley KA, Goldberg NE, Dolan CM, et al. Determining factors that influence parents' perceptions of telehealth provided in a pediatric gastroenterological practice: A quality improvement project. J Pediatr Nurs 2022;62:36-42. [doi: 10.1016/j.pedn.2021.11.023] [Medline: 34894421]
60. Dhahri AA, Iqbal MR, Pardoe H. Agile application of video telemedicine during the COVID-19 pandemic. Cureus 2020 Nov 04;12(11):e11320 [FREE Full text] [doi: 10.7759/cureus.11320] [Medline: 33262918]
61. Dietzen A, Ide W, Pavone L. Telehealth in pediatric rehabilitation medicine: A survey of clinician reported utilization and experience. J Pediatr Rehabil Med 2020;13(3):281-288. [doi: 10.3233/PRM-200762] [Medline: 33252099]
62. Dooley AB, Houssaye NDL, Baum N. Use of telemedicine for sexual medicine patients. Sex Med Rev 2020 Oct;8(4):507-517 [FREE Full text] [doi: 10.1016/j.sxmr.2020.06.001] [Medline: 32739238]
63. Dopelt K, Avni N, Haimov-Sadikov Y, Golan I, Davidovitch N. Telemedicine and eHealth literacy in the era of COVID-19: a cross-sectional study in a peripheral clinic in Israel. Int J Environ Res Public Health 2021 Sep 10;18(18):9556 [FREE Full text] [doi: 10.3390/ijerph18189556] [Medline: 34574480]
64. Board of Directors Meeting - Open (Thursday 27 May 2021). East Kent Hospitals NHS Foundation Trust. 2021. URL: https://www.ekhuft.nhs.uk/about-us/board-of-directors/ [accessed 2024-12-05]

65. Ennis M, Wahl K, Jeong D, Knight K, Renner R, Munro S, et al. The perspective of Canadian health care professionals on abortion service during the COVID-19 pandemic. Fam Pract 2021 Aug 27;38(Suppl 1):i30-i36 [FREE Full text] [doi: 10.1093/fampra/cmab083] [Medline: 34448482]
66. Estudio sobre la consulta de salud virtual y sus beneficios para el sistema sanitario. EsadeGov. 2020 Jul 16. URL: https://tinyurl.com/463ek5dw [accessed 2022-10-26]
67. Esper GJ, Sweeney RL, Winchell E, Duffell JM, Kier SC, Lukens HW, et al. Rapid systemwide implementation of outpatient telehealth in response to the COVID-19 pandemic. J Healthc Manag 2020;65(6):443-452 [FREE Full text] [doi: 10.1097/JHM-D-20-00131] [Medline: 33074968]
68. The organisation of resilient health and social care following the COVID-19 pandemic. European Commission. 2020. URL: https://health.ec.europa.eu/system/files/2020-12/026_health_socialcare_covid19_en_0.pdf [accessed 2022-10-26]
69. Florea M, Lazea C, Gaga R, Sur G, Lotrean L, Puia A, et al. Lights and shadows of the perception of the use of telemedicine by Romanian family doctors during the COVID-19 pandemic. Int J Gen Med 2021;14:1575-1587 [FREE Full text] [doi: 10.2147/IJGM.S309519] [Medline: 33953605]
70. Fluhr JW, Gueguen A, Legoupil D, Brenaut E, Abasq C, Araújo H, et al. Teledermatology in times of COVID-19 confinement: comparing patients' and physicians' satisfaction by the Standardized Brest Teledermatology Questionnaire. Dermatology 2021 Feb 10;237(2):1-6 [FREE Full text] [doi: 10.1159/000514029] [Medline: 33567427]
71. Frank HE, Grumbach NM, Conrad SM, Wheeler J, Wolff J. Mental health services in primary care: Evidence for the feasibility of telehealth during the COVID-19 pandemic. J Affect Disord Rep 2021 Jul;5:100146 [FREE Full text] [doi: 10.1016/j.jadr.2021.100146] [Medline: 33870261]
72. Franzosa E, Gorbenko K, Brody AA, Leff B, Ritchie CS, Kinosian B, et al. "There is something very personal about seeing someone's face": provider perceptions of video visits in home-based primary care during COVID-19. J Appl Gerontol 2021 Nov;40(11):1417-1424 [FREE Full text] [doi: 10.1177/07334648211028393] [Medline: 34210200]
73. Franzosa E, Gorbenko K, Brody AA, Leff B, Ritchie CS, Kinosian B, et al. "At Home, with Care": lessons from New York City home-based primary care practices managing COVID-19. J Am Geriatr Soc 2021 Feb;69(2):300-306 [FREE Full text] [doi: 10.1111/jgs.16952] [Medline: 33179761]
74. Fu L, Lindenmeyer A, Phillimore J, Lessard-Phillips L. Vulnerable migrants' access to healthcare in the early stages of the COVID-19 pandemic in the UK. Public Health 2022 Feb;203:36-42 [FREE Full text] [doi: 10.1016/j.puhe.2021.12.008] [Medline: 35026578]
75. Gold KJ, Laurie AR, Kinney DR, Harmes KM, Serlin DC. Video visits: family physician experiences with uptake during the COVID-19 pandemic. Fam Med 2021 Mar 4;53(3):207-210. [doi: 10.22454/fammed.2021.613099]
76. Gray DP, Sidaway-Lee K, Harding A, Evans P. Reduction in face-to-face GP consultations. Br J Gen Pract 2020 Jul;70(696):328 [FREE Full text] [doi: 10.3399/bjgp20X710849] [Medline: 32586804]
77. Grens H, de Bruin JP, Huppelschoten A, Kremer JAM. Fertility workup with video consultation during the COVID-19 pandemic: pilot quantitative and qualitative study. JMIR Form Res 2022 Feb 07;6(2):e32000. [doi: 10.2196/32000] [Medline: 34936981]
78. Grosman-Dziewiszek P, Wiatrak B, Jęśkowiak I, Szeląg A. Patients' habits and the role of pharmacists and telemedicine as elements of a modern health care system during the COVID-19 pandemic. J Clin Med 2021 Sep 17;10(18):4211 [FREE Full text] [doi: 10.3390/jcm10184211] [Medline: 34575322]
79. Gupta T, Gkiousias V, Bhutta MF. A systematic review of outcomes of remote consultation in ENT. Clin Otolaryngol 2021 Jul;46(4):699-719. [doi: 10.1111/coa.13768] [Medline: 33754458]
80. Hardcastle L, Ogbogu U. Virtual care: Enhancing access or harming care? Healthc Manage Forum 2020 Nov 20;33(6):288-292 [FREE Full text] [doi: 10.1177/0840470420938818] [Medline: 32686506]
81. Hays RD, Skootsky SA. Patient experience with in-person and telehealth visits before and during the COVID-19 pandemic at a large integrated health system in the United States. J Gen Intern Med 2022 Mar;37(4):847-852 [FREE Full text] [doi: 10.1007/s11606-021-07196-4] [Medline: 34982370]
82. Heyck Lee S, Ramondino S, Gallo K, Moist LM. A quantitative and qualitative study on patient and physician perceptions of nephrology telephone consultation during COVID-19. Can J Kidney Health Dis 2022;9:20543581211066720 [FREE Full text] [doi: 10.1177/20543581211066720] [Medline: 35024152]
83. Hofmann G, Hampanda K, Harrison MS, Fasano M, Nacht A, Yeoman M. Virtual prenatal and postpartum care acceptability among maternity care providers. Matern Child Health J 2022 Jul 15;26(7):1401-1408 [FREE Full text] [doi: 10.1007/s10995-022-03412-7] [Medline: 35292887]
84. Holtz BE. Patients perceptions of telemedicine visits before and after the coronavirus disease 2019 pandemic. Telemed J E Health 2021 Jan 01;27(1):107-112 [FREE Full text] [doi: 10.1089/tmj.2020.0168] [Medline: 32614689]
85. Imlach F, McKinlay E, Middleton L, Kennedy J, Pledger M, Russell L, et al. Telehealth consultations in general practice during a pandemic lockdown: survey and interviews on patient experiences and preferences. BMC Fam Pract 2020 Dec 13;21(1):269 [FREE Full text] [doi: 10.1186/s12875-020-01336-1] [Medline: 33308161]
86. Isautier JM, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, et al. People's experiences and satisfaction with telehealth during the COVID-19 pandemic in Australia: cross-sectional survey study. J Med Internet Res 2020 Dec 10;22(12):e24531 [FREE Full text] [doi: 10.2196/24531] [Medline: 33156806]

87. Itamura K, Tang DM, Higgins TS, Rimell FL, Illing EA, Ting JY, et al. Comparison of patient satisfaction between virtual visits during the COVID-19 pandemic and in-person visits pre-pandemic. Ann Otol Rhinol Laryngol 2021 Jul;130(7):810-817 [FREE Full text] [doi: 10.1177/0003489420977766] [Medline: 33251849]
88. Javanparast S, Roeger L, Kwok Y, Reed RL. The experience of Australian general practice patients at high risk of poor health outcomes with telehealth during the COVID-19 pandemic: a qualitative study. BMC Fam Pract 2021 Apr 08;22(1):69 [FREE Full text] [doi: 10.1186/s12875-021-01408-w] [Medline: 33832422]
89. Jhaveri D, Emeto TI, Alele FO, Strom A, Benham H. Use of telemedicine for rheumatology practice in Queensland, Australia: experiences before and during the COVID-19 pandemic. Intern Med J 2022 Oct;52(10):1685-1690. [doi: 10.1111/imj.15706] [Medline: 35112769]
90. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
91. Johnson C, Dupuis JB, Goguen P, Grenier G. Changes to telehealth practices in primary care in New Brunswick (Canada): A comparative study pre and during the COVID-19 pandemic. PLoS One 2021;16(11):e0258839 [FREE Full text] [doi: 10.1371/journal.pone.0258839] [Medline: 34813618]
92. Jonnagaddala J, Godinho MA, Liaw S. From telehealth to virtual primary care in Australia? A Rapid scoping review. Int J Med Inform 2021 Jul;151:104470 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104470] [Medline: 34000481]
93. Kalwani NM, Wang KM, Johnson AN, Deb JD, Gold T, Maddukuri AK, et al. Application of the Quadruple Aim to evaluate the operational impact of a telemedicine program. Healthc (Amst) 2021 Dec;9(4):100593 [FREE Full text] [doi: 10.1016/j.hjdsi.2021.100593] [Medline: 34749227]
94. Kamimura A, Panahi S, Meng H, Sundrud J, Lucero M. Patient satisfaction with telehealth and dxperiences during the COVID-19 pandemic among uninsured free clinic patients. J Patient Exp 2021 Jul 26;8:23743735211033107 [FREE Full text] [doi: 10.1177/23743735211033107] [Medline: 34368425]
95. Karacabeyli D, Shea S, Keidar S, Pinkney S, Bepple K, Edwards D, et al. The Live 5-2-1-0 Toolkit for family physicians: Mixed methods evaluation of a resource to facilitate health promotion in a primary care setting. BC Medical Journal 2020;62(6):1 [FREE Full text]
96. Kato-Lin Y, Thelen ST. Telemedicine for acute conditions during COVID-19: a nationwide survey using crowdsourcing. Telemed J E Health 2021 Jul;27(7):714-723. [doi: 10.1089/tmj.2020.0351] [Medline: 33197368]
97. Khan ZA, Zahoor A, Afzal I, Butt U, Siddiqui AM, Khan Z, et al. Evaluation of patient perception and satisfaction toward the use of telemedicine during pandemic of novel coronavirus in Pakistan. Telemed J E Health 2021 Oct;27(10):1174-1179. [doi: 10.1089/tmj.2020.0343] [Medline: 33449871]
98. Kirby DJ, Fried JW, Buchalter DB, Moses MJ, Hurly ET, Cardone DA, et al. Patient and physician satisfaction with telehealth during the COVID-19 pandemic: sports medicine perspective. Telemed J E Health 2021 Oct;27(10):1151-1159. [doi: 10.1089/tmj.2020.0387] [Medline: 33512302]
99. Kletečka-Pulker M, Völkl-Kernstock S, Fassl A, Klager E, Willschke H, Klomfar S, et al. Telehealth in times of COVID-19: spotlight on Austria. Healthcare (Basel) 2021 Mar 04;9(3):1 [FREE Full text] [doi: 10.3390/healthcare9030280] [Medline: 33806673]
100. Kling SMR, Falco-Walter JJ, Saliba-Gustafsson EA, Garvert DW, Brown-Johnson CG, Miller-Kuhlmann R, et al. Patient and clinician perspectives of new and return ambulatory teleneurology visits. Neurol Clin Pract 2021 Dec;11(6):472-483 [FREE Full text] [doi: 10.1212/CPJ.0000000000001065] [Medline: 34992955]
101. Kludacz-Alessandri M, Walczak R, Hawrysz L, Korneta P. The quality of medical care in the conditions of the COVID-19 pandemic, with particular emphasis on the access to primary healthcare and the effectiveness of treatment in Poland. J Clin Med 2021 Aug 09;10(16):3502 [FREE Full text] [doi: 10.3390/jcm10163502] [Medline: 34441798]
102. Kludacz-Alessandri M, Hawrysz L, Korneta P, Gierszewska G, Pomaranik W, Walczak R. The impact of medical teleconsultations on general practitioner-patient communication during COVID- 19: A case study from Poland. PLoS One 2021;16(7):e0254960 [FREE Full text] [doi: 10.1371/journal.pone.0254960] [Medline: 34270587]
103. Krasovsky T, Silberg T, Barak S, Eisenstein E, Erez N, Feldman I, et al. Transition to multidisciplinary pediatric telerehabilitation during the COVID-19 pandemic: strategy development and implementation. Int J Environ Res Public Health 2021 Feb 04;18(4):1 [FREE Full text] [doi: 10.3390/ijerph18041484] [Medline: 33557395]
104. Kumar S, Kumar A, Kumar M, Kumar A, Arora R, Sehrawat R. Feasibility of telemedicine in maintaining follow-up of orthopaedic patients and their satisfaction: A preliminary study. J Clin Orthop Trauma 2020 Oct;11(Suppl 5):S704-S710 [FREE Full text] [doi: 10.1016/j.jcot.2020.07.026] [Medline: 32837105]
105. Kumar P, Aggarwal M, Dhawan R, Dass J, Kumar G, Sharma V, et al. Tele-medicine services in hematological practice during Covid pandemic: its feasibility and difficulties. Indian J Hematol Blood Transfus 2021 Oct;37(4):528-533 [FREE Full text] [doi: 10.1007/s12288-020-01385-7] [Medline: 33250594]
106. Kurotschka PK, Serafini A, Demontis M, Serafini A, Mereu A, Moro MF, et al. General practitioners' experiences during the first phase of the COVID-19 pandemic in Italy: a critical incident technique study. Front Public Health 2021;9:623904 [FREE Full text] [doi: 10.3389/fpubh.2021.623904] [Medline: 33614587]

107. LeBrun DG, Malfer C, Wilson M, Carroll KM, Wang Ms V, Mayman DJ, et al. Telemedicine in an outpatient arthroplasty setting during the COVID-19 pandemic: early lessons from New York City. HSS J 2021 Feb;17(1):25-30 [FREE Full text] [doi: 10.1177/1556331620972659] [Medline: 33967638]
108. Lewis YD, Elran-Barak R, Grundman-Shem Tov R, Zubery E. The abrupt transition from face-to-face to online treatment for eating disorders: a pilot examination of patients' perspectives during the COVID-19 lockdown. J Eat Disord 2021 Mar 05;9(1):31 [FREE Full text] [doi: 10.1186/s40337-021-00383-y] [Medline: 33673876]
109. Li E, Tsopra R, Jimenez G, Serafini A, Gusso G, Lingner H, et al. General practitioners' perceptions of using virtual primary care during the COVID-19 pandemic: An international cross-sectional survey study. PLOS Digit Health 2022 May;1(5):e0000029 [FREE Full text] [doi: 10.1371/journal.pdig.0000029] [Medline: 36812543]
110. Lopez Guzman Y, Barros Castillo SM, Leal Betancourt MI. Condiciones de vulnerabilidad, sociodemográficas y clínicas de la población con COVID-19 incluida en la estrategia de atención primaria en salud. Mosquera, 2020. Universidad del Rosario. 2021 Oct. URL: https://tinyurl.com/2ckxdpnu [accessed 2022-10-26]
111. Maher DP, Hess D, Edwards C, Allen L. Changes in patient satisfaction scores during the early COVID-19 pandemic. J Patient Exp 2021;8:23743735211034610 [FREE Full text] [doi: 10.1177/23743735211034610] [Medline: 34377775]
112. Malliaras P, Merolli M, Williams CM, Caneiro JP, Haines T, Barton C. 'It's not hands-on therapy, so it's very limited': Telehealth use and views among allied health clinicians during the coronavirus pandemic. Musculoskelet Sci Pract 2021 Apr;52:102340 [FREE Full text] [doi: 10.1016/j.msksp.2021.102340] [Medline: 33571900]
113. Manski-Nankervis J, Davidson S, Hiscock H, Hallinan C, Ride J, Lingam V, et al. Primary care consumers' experiences and opinions of a telehealth consultation delivered via video during the COVID-19 pandemic. Aust J Prim Health 2022 Jun;28(3):224-231. [doi: 10.1071/PY21193] [Medline: 35287793]
114. Massaroni V, Delle Donne V, Ciccarelli N, Ciccullo A, Borghetti A, Faliero D, et al. Use of telehealth for HIV care in Italy: Are doctors and patients on the same page? A cross-sectional study. Int J Med Inform 2021 Dec;156:104616. [doi: 10.1016/j.ijmedinf.2021.104616] [Medline: 34695728]
115. Mathew T, Lee PC, Ianno DJ, Benson J. Telehealth and Australian general practice in 2020: A survey exploring patients' perspectives in the Adelaide Hills. Aust J Gen Pract 2021 Oct;50(10):754-759 [FREE Full text] [doi: 10.31128/AJGP-11-20-5727] [Medline: 34590090]
116. Mbunge E, Batani J, Gaobotse G, Muchemwa B. Virtual healthcare services and digital health technologies deployed during coronavirus disease 2019 (COVID-19) pandemic in South Africa: a systematic review. Glob Health J 2022 Jun;6(2):102-113 [FREE Full text] [doi: 10.1016/j.glohj.2022.03.001] [Medline: 35282399]
117. Menhadji P, Patel R, Asimakopoulou K, Quinn B, Khoshkhounejad G, Pasha P, et al. Patients' and dentists' perceptions of tele-dentistry at the time of COVID-19. A questionnaire-based study. J Dent 2021 Oct;113:103782 [FREE Full text] [doi: 10.1016/j.jdent.2021.103782] [Medline: 34400252]
118. Miller MJ, Pak SS, Keller DR, Barnes DE. Evaluation of pragmatic telehealth physical therapy implementation during the COVID-19 pandemic. Phys Ther 2021 Jan 04;101(1):1 [FREE Full text] [doi: 10.1093/ptj/pzaa193] [Medline: 33284318]
119. Mills KT, Peacock E, Chen J, Zimmerman A, He H, Cyprian A, et al. Experiences and beliefs of low-income patients with hypertension in Louisiana and Mississippi during the COVID-19 pandemic. J Am Heart Assoc 2021 Feb 02;10(3):e018510 [FREE Full text] [doi: 10.1161/JAHA.120.018510] [Medline: 33267723]
120. Moore A, Chavez C, Fisher MP. Factors enhancing trust in electronic communication among patients from an internal medicine clinic: qualitative results of the RECEPT study. J Gen Intern Med 2022 Sep 19;37(12):3121-3127 [FREE Full text] [doi: 10.1007/s11606-021-07345-9] [Medline: 35048293]
121. Mostafa PIN, Hegazy AA. Dermatological consultations in the COVID-19 era: is teledermatology the key to social distancing? An Egyptian experience. J Dermatolog Treat 2022 Mar;33(2):910-915. [doi: 10.1080/09546634.2020.1789046] [Medline: 32602763]
122. Gorgojo Martínez JJ, Murillo AZ, Rubio Herrera M, Lesmes IB. TELECONSULTA en Endocrinología y Nutrición en tiempos de la pandemia COVID-19 y más allá. Sociedad Española de Endocrinología y Nutrición. URL: https://tinyurl.com/2jhyx2es [accessed 2024-11-06]
123. Murren-Boezem J, Solo-Josephson P, Zettler-Greeley CM. On-demand, virtual health care during COVID-19: clinician redeployment and telemedicine utilization in a children's health system. Telemed J E Health 2021 Oct;27(10):1111-1116. [doi: 10.1089/tmj.2020.0461] [Medline: 33393879]
124. Mustafa SS, Vadamalai K, Ramsey A. Patient satisfaction with in-person, video, and telephone allergy/immunology evaluations during the COVID-19 pandemic. J Allergy Clin Immunol Pract 2021 May;9(5):1858-1863 [FREE Full text] [doi: 10.1016/j.jaip.2021.01.036] [Medline: 33548516]
125. Taskforce on Telehealth Policy (TTP): Findings and Recommendations. National Committee for Quality Assurance (NCQA). URL: https://tinyurl.com/4j67w4ju [accessed 2022-10-26]
126. Nguyen OT, Watson AK, Motwani K, Warpinski C, McDilda K, Leon C, et al. Patient-level factors associated with utilization of telemedicine services from a free clinic during COVID-19. Telemed J E Health 2022 Apr;28(4):526-534. [doi: 10.1089/tmj.2021.0102] [Medline: 34255572]

127. Olayiwola JN, Magaña C, Harmon A, Nair S, Esposito E, Harsh C, et al. Telehealth as a bright spot of the COVID-19 pandemic: recommendations from the virtual frontlines ("Frontweb"). JMIR Public Health Surveill 2020 Jun 25;6(2):e19045 [FREE Full text] [doi: 10.2196/19045] [Medline: 32479413]
128. O'Neill C, Matias MV, Peixoto VR, O'Neill H, Aguiar P, Castela E. [Teleconsultation in the Portuguese National Health Service during the COVID-19 pandemic: a survey of physicians' views and future implications]. Acta Med Port 2022 Jun 01;35(6):455-467 [FREE Full text] [doi: 10.20344/amp.16724] [Medline: 35300783]
129. La Atención Primaria Rural en España lanza un grito de socorro para su supervivencia. Médicos y Pacientes. URL: https://tinyurl.com/3jna2w96 [accessed 2022-11-05]
130. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589 [FREE Full text] [doi: 10.2196/28589] [Medline: 33822736]
131. Parker K, Chia M. Patient and clinician satisfaction with video consultations in dentistry - part one: patient satisfaction. Br Dent J 2021 May 27:1-6 [FREE Full text] [doi: 10.1038/s41415-021-3007-y] [Medline: 34045675]
132. Parsons MW, Gardner MM, Sherman JC, Pasquariello K, Grieco JA, Kay CD, et al. Feasibility and acceptance of direct-to-home tele-neuropsychology services during the COVID-19 pandemic. J Int Neuropsychol Soc 2022 Feb;28(2):210-215. [doi: 10.1017/S1355617721000436] [Medline: 33952375]
133. Patel S, Douglas-Moore J. A reflection on an adapted approach from face-to-face to telephone consultations in our Urology Outpatient Department during the COVID-19 pandemic - a pathway for change to future practice? BJU Int 2020 Sep 04;126(3):339-341 [FREE Full text] [doi: 10.1111/bju.15119] [Medline: 32469096]
134. Written evidence submitted by Philips UKI (HSC0978). UK Parliament. 2021 Mar. URL: https://committees.parliament.uk/writtenevidence/25169/pdf/ [accessed 2022-10-18]
135. Rego N, Pereira HS, Crispim J. Perceptions of patients and physicians on teleconsultation at home for diabetes mellitus: survey study. JMIR Hum Factors 2021 Nov 23;8(4):e27873 [FREE Full text] [doi: 10.2196/27873] [Medline: 34817394]
136. Reicher S, Sela T, Toren O. Using telemedicine during the COVID-19 pandemic: attitudes of adult health care consumers in Israel. Front Public Health 2021;9:653553 [FREE Full text] [doi: 10.3389/fpubh.2021.653553] [Medline: 34079784]
137. Reynolds-Wright JJ, Boydell N, Cameron S, Harden J. A qualitative study of abortion care providers' perspectives on telemedicine medical abortion provision in the context of COVID-19. BMJ Sex Reprod Health 2022 Jul;48(3):199-204 [FREE Full text] [doi: 10.1136/bmjsrh-2021-201309] [Medline: 34848554]
138. Riedel L, Uscher-Pines L, Mehrotra A, Busch AB, Barnett ML, Raja P, et al. Use of telemedicine for opioid use disorder treatment - Perceptions and experiences of opioid use disorder clinicians. Drug Alcohol Depend 2021 Nov 01;228:108999 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108999] [Medline: 34517225]
139. Rosler G. Pediatric telehealth experiences: myths and truths about video visits from a parent. J Patient Exp 2020 Dec;7(6):836-838 [FREE Full text] [doi: 10.1177/2374373520932724] [Medline: 33457506]
140. General practice in the post Covid world: challenges and opportunities for general practice. Royal College of General Practitioners. URL: https://tinyurl.com/4kkaxtxy [accessed 2024-11-06]
141. Focus: artificial intelligence and technology in family medicine. Family Doctor. 2021. URL: https://www.nysafp.org/wp-content/uploads/2024/05/Winter-2021.pdf [accessed 2024-11-06]
142. Saiyed S, Nguyen A, Singh R. Physician perspective and key satisfaction indicators with rapid telehealth adoption during the coronavirus disease 2019 pandemic. Telemed J E Health 2021 Nov;27(11):1225-1234. [doi: 10.1089/tmj.2020.0492] [Medline: 33513045]
143. Scherer J, Back DA, Thienemann F, Kaufmann E, Neuhaus V, Willy C, et al. The effect of Covid-19 on the willingness to use video consultations among orthopedic and trauma outpatients: a multi-center survey in 1400 outpatients. Eur J Trauma Emerg Surg 2022 Jun;48(3):2199-2206 [FREE Full text] [doi: 10.1007/s00068-021-01774-1] [Medline: 34427693]
144. Schweiberger K, Hoberman A, Iagnemma J, Schoemer P, Squire J, Taormina J, et al. Practice-level variation in telemedicine use in a pediatric primary care network during the COVID-19 pandemic: retrospective analysis and survey study. J Med Internet Res 2020 Dec 18;22(12):e24345 [FREE Full text] [doi: 10.2196/24345] [Medline: 33290244]
145. Shachak A, Alkureishi MA. Virtual care: a 'Zoombie' apocalypse? J Am Med Inform Assoc 2020 Nov 01;27(11):1813-1815 [FREE Full text] [doi: 10.1093/jamia/ocaa185] [Medline: 32940711]
146. Sharma AE, Khoong EC, Sierra M, Rivadeneira NA, Nijagal MA, Su G, et al. System-level factors associated with telephone and video visit use: survey of safety-net clinicians during the early phase of the COVID-19 pandemic. JMIR Form Res 2022 Mar 10;6(3):e34088 [FREE Full text] [doi: 10.2196/34088] [Medline: 35148271]
147. Skelly JR, O'Connor T. Guidelines for the use of the Attend Anywhere Platform for Telecommunications within the pain service. Ir Med J 2021 Aug 19;114(7):403. [Medline: 34520346]
148. Srinivasan M, Asch S, Vilendrer S, Thomas SC, Bajra R, Barman L, et al. Qualitative assessment of rapid system transformation to primary care video visits at an academic medical center. Annals of Internal Medicine 2020 Oct 06;173(7):527-535. [doi: 10.7326/m20-1814]
149. Staffieri SE, Mathew AA, Sheth SJ, Ruddle JB, Elder JE. Parent satisfaction and acceptability of telehealth consultations in pediatric ophthalmology: initial experience during the COVID-19 pandemic. J AAPOS 2021 Apr;25(2):104-107. [doi: 10.1016/j.jaapos.2020.12.003] [Medline: 33689911]

150. Sugarman DE, Horvitz LE, Greenfield SF, Busch AB. Clinicians' perceptions of rapid scale-up of telehealth services in outpatient mental health treatment. Telemed J E Health 2021 Dec;27(12):1399-1408 [FREE Full text] [doi: 10.1089/tmj.2020.0481] [Medline: 33600272]
151. Hamui-Sutton L, Halabe-Cherem J. Videoconsultas: alcances y limitaciones. Medicina Interna de México 2021;37(2):261-266 [FREE Full text]
152. Swaminathan R, Mughal Z, Phillips D. Telephone Consultation in Otorhinolaryngology During the Coronavirus Disease 2019 Pandemic: a Cross-sectional Analysis of Effectiveness and Satisfaction for Patients and Clinicians. SN Compr Clin Med 2022;4(1):36 [FREE Full text] [doi: 10.1007/s42399-022-01119-y] [Medline: 35036851]
153. Tai-Seale M, Rosen R, Ruo B, Hogarth M, Longhurst CA, Lander L, et al. Implementation of patient engagement tools in electronic health records to enhance patient-centered communication: protocol for feasibility evaluation and preliminary results. JMIR Res Protoc 2021 Aug 26;10(8):e30431 [FREE Full text] [doi: 10.2196/30431] [Medline: 34435960]
154. Taylor P, Berg C, Thompson J, Dean K, Yuan T, Nallamshetty S, et al. Effective access to care in a crisis period: hypertension control during the COVID-19 pandemic by telemedicine. Mayo Clin Proc Innov Qual Outcomes 2022 Feb;6(1):19-26 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2021.11.006] [Medline: 34805763]
155. Tenforde AS, Iaccarino MA, Borgstrom H, Hefner JE, Silver J, Ahmed M, et al. Telemedicine during COVID-19 for outpatient sports and musculoskeletal medicine physicians. PM R 2020 Sep;12(9):926-932 [FREE Full text] [doi: 10.1002/pmrj.12422] [Medline: 32424977]
156. Teng T, Sareidaki DE, Chemaly N, Bar C, Coste-Zeitoun D, Kuchenbuch M, et al. Physician and patient satisfaction with the switch to remote outpatient encounters in epilepsy clinics during the Covid-19 pandemic. Seizure 2021 Oct;91:60-65 [FREE Full text] [doi: 10.1016/j.seizure.2021.05.013] [Medline: 34098318]
157. Thiele T, Beider S, Kühl H, Mielke G, Holz A, Hirsch S, et al. [Care of rheumatology patients during the lockdown in early 2020 : Telemedicine, delegation, patient satisfaction and vaccination behavior]. Z Rheumatol 2022 Mar;81(2):157-163 [FREE Full text] [doi: 10.1007/s00393-021-01005-3] [Medline: 33974131]
158. Tornero-Molina J, Sánchez-Alonso F, Fernández-Prada M, Bris-Ochaita M, Sifuentes-Giraldo A, Vidal-Fuentes J. Tele-rheumatology during the COVID-19 pandemic. Reumatol Clin (Engl Ed) 2020 Oct 29;18(3):157-163 [FREE Full text] [doi: 10.1016/j.reuma.2020.10.004] [Medline: 33214110]
159. Tozour JN, Bandremer S, Patberg E, Zavala J, Akerman M, Chavez M, et al. Application of telemedicine video visits in a maternal-fetal medicine practice at the epicenter of the COVID-19 pandemic. Am J Obstet Gynecol MFM 2021 Nov;3(6):100469 [FREE Full text] [doi: 10.1016/j.ajogmf.2021.100469] [Medline: 34450341]
160. Triantafillou V, Layfield E, Prasad A, Deng J, Shanti RM, Newman JG, et al. Patient perceptions of head and neck ambulatory telemedicine visits: a qualitative study. Otolaryngol Head Neck Surg 2021 May 14;164(5):923-931. [doi: 10.1177/0194599820943523] [Medline: 32662749]
161. Troncone A, Cascella C, Chianese A, Zanfardino A, Casaburo F, Piscopo A, et al. Doctor-patient relationship in synchronous/real-time video-consultations and in-person visits: an investigation of the perceptions of young people with type 1 diabetes and their parents during the COVID-19 pandemic. Int J Behav Med 2022 Oct;29(5):638-647 [FREE Full text] [doi: 10.1007/s12529-021-10047-5] [Medline: 35076861]
162. Tulupova E, Dobiášová K, Kopsa Těšinová J. Patient experience with telemedicine in primary care during the COVID pandemic: a qualitative study. European Journal of Public Health 2021;31(Supplement_3):ckab165.162. [doi: 10.1093/eurpub/ckab165.162]
163. Utley LM, Manchala GS, Phillips MJ, Doshi CP, Szatalowicz VL, Boozer JR. Bridging the telemedicine gap among seniors during the COVID-19 pandemic. J Patient Exp 2021;8:23743735211014036 [FREE Full text] [doi: 10.1177/23743735211014036] [Medline: 34179442]
164. Verduzco-Gutierrez M, Bean AC, Tenforde AS, Tapia RN, Silver JK. How to conduct an outpatient telemedicine rehabilitation or prehabilitation visit. PM R 2020 Jul 07;12(7):714-720. [doi: 10.1002/pmrj.12380] [Medline: 32297458]
165. Verma P, Kerrison R. Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. BJGP Open 2022 Jun;6(2):1 [FREE Full text] [doi: 10.3399/BJGPO.2021.0192] [Medline: 35031558]
166. Villalobos J, Calvanese N. [The impact of COVID-19 pandemic on doctor-patient relationship]. Rev Med Chil 2021 Jul;149(7):1070-1074 [FREE Full text] [doi: 10.4067/s0034-98872021000701070] [Medline: 34751309]
167. Volandes AE, Zupanc SN, Paasche-Orlow MK, Lakin JR, Chang Y, Burns EA, et al. Association of an advance care planning video and communication intervention with documentation of advance care planning among older adults: a nonrandomized controlled trial. JAMA Netw Open 2022 Feb 01;5(2):e220354 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.0354] [Medline: 35201306]
168. Vongsachang H, Lagstein O, Boland MV, Repka MX, Kraus CL, Collins ME. Telemedicine utilization by pediatric ophthalmologists during the COVID-19 pandemic. J AAPOS 2021 Oct;25(5):293-295.e1 [FREE Full text] [doi: 10.1016/j.jaapos.2021.05.018] [Medline: 34600105]
169. Webber EC, McMillen BD, Willis DR. Health care disparities and access to video visits before and after the COVID-19 pandemic: findings from a patient survey in primary care. Telemed J E Health 2022 May;28(5):712-719. [doi: 10.1089/tmj.2021.0126] [Medline: 34449270]

170. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
171. Wilson G, Windner Z, Bidwell S, Currie O, Dowell A, Halim AA, et al. 'Here to stay': changes to prescribing medication in general practice during the COVID-19 pandemic in New Zealand. J. Prim. Health Care 2021 Aug 13;13(3):222-230. [doi: 10.1071/hc21035]
172. Hoda R. Socio-technical grounded theory for software engineering. IIEEE Trans. Software Eng 2022 Oct 1;48(10):3808-3832. [doi: 10.1109/tse.2021.3106280]
173. Chitungo I, Mhango M, Dzobo M, Denhere K, Chimene M, Musuka G, et al. Towards virtual doctor consultations: A call for the scale-up of telemedicine in sub-Saharan Africa during COVID-19 lockdowns and beyond. Smart Health (Amst) 2021 Jul;21:100207 [FREE Full text] [doi: 10.1016/j.smhl.2021.100207] [Medline: 36570120]
174. Nittari G, Khuman R, Baldoni S, Pallotta G, Battineni G, Sirignano A, et al. Telemedicine practice: review of the current ethical and legal challenges. Telemed J E Health 2020 Dec;26(12):1427-1437 [FREE Full text] [doi: 10.1089/tmj.2019.0158] [Medline: 32049608]
175. de Oliveira Andrade A, Soares AB, de Andrade Palis A, Cabral AM, Barreto CGL, de Souza DB, et al. On the use of telemedicine in the context of COVID-19: legal aspects and a systematic review of technology. Res. Biomed. Eng 2021 Mar 13;38(1):209-227. [doi: 10.1007/s42600-021-00133-8]
176. Ftouni R, AlJardali B, Hamdanieh M, Ftouni L, Salem N. Challenges of telemedicine during the COVID-19 pandemic: a systematic review. BMC Med Inform Decis Mak 2022 Aug 03;22(1):207 [FREE Full text] [doi: 10.1186/s12911-022-01952-0] [Medline: 35922817]
177. Pogorzelska K, Chlabicz S. Patient satisfaction with telemedicine during the COVID-19 pandemic-a systematic review. Int J Environ Res Public Health 2022 May 17;19(10):6113 [FREE Full text] [doi: 10.3390/ijerph19106113] [Medline: 35627650]
178. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
179. Silva A, Sindico S, Carneiro A, Henrique S, Fernandes A, Gomes J, et al. COVID-19 remote consultation services and population in health inequity-concentrating territories: a scoping review. Telemed J E Health 2021 Aug;27(8):881-897 [FREE Full text] [doi: 10.1089/tmj.2021.0145] [Medline: 34232749]
180. Boland J, Banks S, Krabbe R, Lawrence S, Murray T, Henning T, et al. A COVID-19-era rapid review: using Zoom and Skype for qualitative group research. Public Health Res Pract 2022 Jun 15;32(2):1 [FREE Full text] [doi: 10.17061/phrp31232112] [Medline: 35702751]
181. Andrews E, Berghofer K, Long J, Prescott A, Caboral-Stevens M. Satisfaction with the use of telehealth during COVID-19: An integrative review. Int J Nurs Stud Adv 2020 Nov;2:100008 [FREE Full text] [doi: 10.1016/j.ijnsa.2020.100008] [Medline: 33083791]
182. Beattie M, Lauder W, Atherton I, Murphy DJ. Instruments to measure patient experience of health care quality in hospitals: a systematic review protocol. Syst Rev 2014 Jan 04;3(1):4 [FREE Full text] [doi: 10.1186/2046-4053-3-4] [Medline: 24387141]
183. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022 Dec;49(4):517-530 [FREE Full text] [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
184. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Soc Sci Med 1998 Nov;47(9):1351-1359. [doi: 10.1016/s0277-9536(98)00213-5] [Medline: 9783878]
185. Williams S, Weinman J, Dale J. Doctor-patient communication and patient satisfaction: a review. Fam Pract 1998 Oct;15(5):480-492. [doi: 10.1093/fampra/15.5.480] [Medline: 9848436]
186. Wolf JA, Niederhauser V, Marshburn D, LaVela SL. Defining patient experience. Patient Experience Journal 2014;1(1):7-19. [doi: 10.35680/2372-0247.1004]
187. Garcia R, Adelakun O. A conceptual framework and pilot study for examining telemedicine satisfaction research. J Med Syst 2019 Jan 25;43(3):51. [doi: 10.1007/s10916-019-1161-4] [Medline: 30684065]
188. Wang T, Giunti G, Melles M, Goossens R. Digital patient experience: umbrella systematic review. J Med Internet Res 2022 Aug 04;24(8):e37952 [FREE Full text] [doi: 10.2196/37952] [Medline: 35925651]
189. Greenhalgh T, Ladds E, Hughes G, Moore L, Wherton J, Shaw SE, et al. Why do GPs rarely do video consultations? qualitative study in UK general practice. Br J Gen Pract 2022 Feb 16;72(718):e351-e360. [doi: 10.3399/bjgp.2021.0658]
190. Mohtar LN, Badr NG. Digital Bridge or Tradeoff: Telehealth Adoption and Healthcare Service Quality. A Scoping Review. In: Za S, Winter R, Lazazzara A, editors. Sustainable Digital Transformation. Lecture Notes in Information Systems and Organisation, vol 59. Cham, Switzerland: Springer International Publishing; 2023:253-268.
191. Lauffenburger JC, Khatib R, Siddiqi A, Albert MA, Keller PA, Samal L, et al. Reducing ethnic and racial disparities by improving undertreatment, control, and engagement in blood pressure management with health information technology (REDUCE-BP) hybrid effectiveness-implementation pragmatic trial: Rationale and design. Am Heart J 2023 Jan;255:12-21 [FREE Full text] [doi: 10.1016/j.ahj.2022.10.003] [Medline: 36220355]

192. Jaworski BK, Webb Hooper M, Aklin WM, Jean-Francois B, Elwood WN, Belis D, et al. Advancing digital health equity: Directions for behavioral and social science research. Transl Behav Med 2023 Apr 03;13(3):132-139 [FREE Full text] [doi: 10.1093/tbm/ibac088] [Medline: 36318232]
193. Kelly M, Fullen B, Martin D, Bradley C, O'Mahony B, McVeigh J. Design and development of an eHealth intervention to support self-management in people with musculoskeletal disorders - 'eHealth: It's TIME': a study protocol. HRB Open Res 2022 Nov 14;5:73. [doi: 10.12688/hrbopenres.13611.1]
194. Caillaud C, Lazarevic N, Pizzuti C, Rosic G, Boehm C, Williams K. Digital Health as a silver lining: exploring pregnancy digital health usage amidst the COVID-19 pandemic. Digitial Health Week 2023 Dec 19;2(1):1 [FREE Full text] [doi: 10.21203/rs.3.rs-2115070/v1]
195. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
196. Ragnedda M, Ruiu ML, Addeo F. The self-reinforcing effect of digital and social exclusion: The inequality loop. Telematics and Informatics 2022 Aug;72:101852. [doi: 10.1016/j.tele.2022.101852]
197. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
198. Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. J Clin Epidemiol 2019 Jul;111:127-134. [doi: 10.1016/j.jclinepi.2019.03.004] [Medline: 30951800]
199. Peters M, Marnie C, Tricco A, Pollock D, Munn Z, Alexander L. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020:18-26. [doi: 10.11124/jbies-20-00167]
200. Valdes D. Global evidence on the rapid adoption of telemedicine in primary care in the COVID-19 pandemic. OSF. 2023 Jan 02. URL: https://osf.io/4z5ut/ [accessed 2024-11-06]
201. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8(1):117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]

Abbreviations

JBI: Joanna Briggs Institute
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
PCC: Population, Concept, Context
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 27.12.22; peer-reviewed by S Vanderhout, A Frank, E Hadley; comments to author 06.04.23; revised version received 29.07.23; accepted 25.07.24; published 28.02.25.

Please cite as:
Valdes D, Shanker A, Hijazi G, Mensah DO, Bockarie T, Lazar I, Ibrahim SA, Zolfagharinia H, Procter R, Spencer R, Dale J, Paule A, Medlin LJ, Tharuvara Kallottil K
Global Evidence on the Sustainability of Telemedicine in Outpatient and Primary Care During the First 2 Years of the COVID-19 Pandemic: Scoping Review Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) Framework
Interact J Med Res 2025;14:e45367
URL: https://www.i-jmr.org/2025/1/e45367
doi:10.2196/45367
PMID:40053716

Review

Therapeutic Guidelines for the Self-Management of Major Depressive Disorder: Scoping Review

Priscila de Campos Tibúrcio[1], MS; Priscila Maria Marcheti[2], PhD; Daniela Miori Pascon[3], PhD; Marco Antônio Montebello Junior[4], MBA; Maria Alzete de Lima[5], PhD; Carla Sílvia Fernandes[6], PhD; Célia Samarina Vilaça de Brito Santos[6], PhD; Maria do Perpétuo Socorro de Sousa Nóbrega[1], PhD

[1]School of Nursing, University of São Paulo, São Paulo, Brazil

[2]School of Nursing, Federal University of Mato Grosso do Sul, Campo Grande, Brazil

[3]School of Nursing, Pontifical Catholic University of São Paulo, Sorocaba, Brazil

[4]Sorocaba Engineering Department, FACENS University Center, Sorocaba, Brazil

[5]School of Nursing, Federal University of Rio Grande do Norte, Rio Grande do Norte, Natal, Brazil

[6]Nursing Department, Porto School of Nursing, Porto, Portugal

Corresponding Author:
Priscila de Campos Tibúrcio, MS
School of Nursing
University of São Paulo
Av. Dr. Enéas de Carvalho Aguiar, 419
São Paulo, 05403-000
Brazil
Phone: 55 19997889041
Email: priscila.campostib13@gmail.com

Abstract

Background: Major depressive disorder contributes to the global burden of mental illness. Therapeutic guidelines promote treatment self-management and support caregivers and family members in this process.

Objective: We aimed to identify therapeutic guidelines for the symptoms of major depressive disorder.

Methods: This scoping review followed the assumptions established by the Joanna Briggs Institute and the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) protocol, carried out in 12 databases (LILACS, PubMed, SciELO, Scopus, Web of Science, b-on, BDENF, AgeLine, Cochrane, BVS, IBECS, and CINAHL) and 5 secondary gray literature sources (Google Scholar, Global ETD Search, EBSCO Open Dissertations, CAPES Catalog of Theses and Dissertations, and the Digital Library of Theses and Dissertations of the University of Sao Paulo). The eligibility criteria were based on the population, concept, and context framework: people diagnosed with major depressive disorder aged >18 years (population), therapeutic guidelines for self-management of major depressive disorder symptoms (concept), and symptoms of major depressive disorder (context). Data collection was carried out from March to July 2022 and updated in June 2024. The included studies were experimental, quasi-experimental, analytical observational, descriptive observational, qualitative, or quantitative studies; systematic reviews and meta-analyses; and scoping and literature reviews published in full without time restrictions in English, Spanish, or Portuguese. All the information, as well as the studies captured, was stored in a Microsoft Excel spreadsheet using Rayyan and the *JBI Manual for Evidence Synthesis*. The titles, abstracts, and full texts were carefully read and classified, extracting the results. After review by 2 independent researchers, 62 studies were selected. The results are presented descriptively, including characterization of the studies and mapping and categorization of groups and subgroups of therapeutic guidelines for self-management of major depressive disorder.

Results: In total, 62 studies published between 2011 and 2023 were included, where 44 (71%) came from indexed data sources and 18 (29%) were gray literature indexed on Google Scholar (13/62, 21%), doctoral theses (3/62, 5%), and master's dissertations (2/62, 3%). Among the therapeutic guidelines identified, mapped, and categorized, 7 major groups were identified for self-management: psychotherapy (32/62, 52%), adoption of healthy habits (25/62, 40%), integrative and complementary practices (17/62, 27%), relaxation techniques (9/62, 14%), consultation with a health professional (14/62, 22%), pharmacological therapy (9/62, 14%), and leisure or pleasurable activities (4/62, 6%).

Conclusions: It was possible to identify therapeutic guidelines to promote self-management of major depressive disorder in the adult population. Therapeutic guidance is an important resource for patients, their families, and the community, making patients the protagonists of their own health. For health professionals, therapeutic guidelines become tools that help develop skills and competencies for care among patients, thus ensuring their ability to self-manage major depressive disorder.

(Interact J Med Res 2025;14:e63959) doi:10.2196/63959

KEYWORDS

major depressive disorder; nursing; revision; self-management; symptoms; PRISMA

Introduction

Background

Major depressive disorder is one of the most prevalent mental disorders in the world, as well as one of the clinical conditions that contributes most to the global burden of mental illnesses [1]. The number of people presenting with symptoms suggestive of this disorder has shown a significant increase at an alarming and worrying rate, and it is estimated that it will be the world's first major disabling public health problem by 2030 [2].

It is a classic mental disorder that involves evident alterations in affect, cognition, and neurovegetative functions, affecting approximately 300 million people worldwide [2]. It is estimated that 3.8% of cases lead to functional incapacity and damage to physical and mental health, as well as professional losses and considerable morbidity and mortality due to suicide or association with other illnesses [3].

In addition to these impacts, it is difficult to diagnose major depressive disorder accurately and quickly as its classification and assessment are based on clinical findings and the patient's history [2]. Therapies for the treatment of major depressive disorder, such as psychobiosocial therapies, psychotherapy, and pharmacotherapy, have advanced and are used according to the severity of the symptoms and potential adverse events [4]. Considering that mental health care has undergone important transformations, such as the creation of an out-of-hospital network made up of substitute services that propose the rescue of the singularity or subjectivity of the person in psychological distress, respecting their integrality and existence [5], promoting their autonomy in the management of the illness or treatment means encouraging empowerment, the ability to make choices, and taking responsibility for themselves as a citizen and social being [6].

This knowledge and skills are important tools to prevent complications, control symptoms, self-manage treatment, and even avoid recurrences and hospitalizations [7]. Due to the complexity of the disease, some adaptive actions can be taken, including controlling destabilizing factors, preserving self-confidence, re-establishing important relationships, attempting to regain functioning, managing symptoms, negotiating the care environment, and maintaining satisfying relationships [7,8].

Society has been moving toward the construction and adoption of technological innovations using smartphones and computer applications. These resources can be of great value in the process of promoting people's skills in the mental health field [8]. In this sense, based on the results of this scoping review, a systematic set of therapeutic guidelines was developed based on clinical practice and scientific evidence. This set of therapeutic guidelines is capable of supporting a computational application, generating knowledge in the area of mental health, creating a connection with the patient, and supporting self-management of one's health remotely [8,9].

These therapeutic guidelines can be adopted with the aim of promoting and supporting self-management in the daily treatment of the patient and supporting the caregiver or family members in performing techniques safely and appropriately, completing therapeutic sessions, maintaining functionality, or preventing possible complications of the disorder [9]. It should be emphasized that, although it is important to encourage self-management by the individual and the family, it is also essential that the care provided by health professionals, including nurses, assists in the promotion and reorganization of the self, breaks and overcomes relationships of dependence, and levels the acquisition or development of knowledge and skills for self-management [7,9].

Given the extent of the symptomatology and the particularities of major depressive disorder, a scoping review allows for an exhaustive analysis of the studies available so that the patient has access to information that will allow them to develop attitudes, knowledge, and skills to self-manage based on the best scientific evidence.

Objectives

Similarly, a set of therapeutic guidelines will constitute an instrument for the clinical practice of nurses and of the multidisciplinary team as the deepening of specific skills will allow for the provision of individualized care that is more suited to the needs of the person with major depressive disorder. Therefore, this scoping review was carried out with the aim of identifying therapeutic guidelines in scientific production for the self-management of major depressive disorder symptoms.

Methods

Overview

A scoping review is used to map key concepts, examine existing evidence before conducting a systematic review, and clarify and define conceptual boundaries. This scoping review was prepared in accordance with the methodological proposal of the Joanna Briggs Institute (JBI) in 2020, supported by the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [10] and PRISMA extension for Scoping Reviews (PRISMA-ScR) [11] guidelines based on the following steps: formulation of the research question, identification of

relevant studies, selection of studies, extraction and analysis of data, and synthesis and construction of the report. The PRISMA-ScR checklist is available in Multimedia Appendix 1. This review protocol was registered in the Open Science Framework platform [12] on November 11, 2023, and was modified on February 19, 2025. There was no deviation from the protocol registered on the Open Science Framework platform.

Research Question

To formulate the guiding research question and guide data collection, the population, concept, and context (PCC) framework was used. The PCC strategy was adopted to drive the research question of this scoping review. In this study, *population* refers to people diagnosed with major depressive disorder aged >18 years, *concept* refers to therapeutic guidelines for self-management of major depressive disorder symptoms, and *context* refers to major depressive disorder symptoms (eg, changes in weight, appetite, and sleep; feelings of guilt; irritability or bad mood; anhedonia; fatigue; and low self-esteem).

Therefore, the following guiding question was defined: "What therapeutic guidelines support self-management of changes in weight, appetite, sleep, feelings of guilt, irritability or bad mood, anhedonia, fatigue and low self-esteem?"

Study Design and Eligibility Criteria

The data search was carried out in 12 information sources (LILACS, PubMed, SciELO, Scopus, Web of Science, b-on, BDENF, AgeLine, Cochrane, BVS, IBECS, and CINAHL) and 5 secondary information sources of gray literature (Google Scholar, Global ETD Search, EBSCO Open Dissertations, CAPES Catalog of Theses and Dissertations, and the Digital Library of Theses and Dissertations of the University of Sao Paulo).

The search strategy was developed using the controlled and noncontrolled descriptors obtained in the initial search plus the Boolean operators "OR," "NOT," and "AND," as well as keywords found in the Health Sciences Descriptors and Medical Subject Headings of the US National Library of Medicine combined with each other according to each database. The strategies used and how they were combined are available in Multimedia Appendix 2.

The search followed five distinct phases according to the JBI methodology, and a team of 5 researchers was assembled for this scoping review: (1) initial search in the selected databases to identify articles on the topic and, from there, select words and indexing terms contained in these publications to develop the full search strategy; (2) use of the keywords and indexing terms identified to search all the databases included; (3) definition of the study design and eligibility criteria, including the determination of the type of study, target population, interventions and outcomes analyzed. In addition, inclusion and exclusion criteria were established for the selection of studies; (4) identification and selection of studies, including the systematic search in databases and sources of gray literature, application of filters and search strategies to ensure the identification of relevant studies and screening of relevant articles (reading of titles, abstracts and, when necessary, the full text); and (5) data extraction and analysis, including the collection of information from the selected studies, the assessment of methodological quality and the synthesis of findings.

The eligibility criteria are described in Textbox 1, and the studies found were refined based on the PCC acronym, type of study, year of publication, language, and results. The search was carried out from March 2022 to July 2022 and updated in June 2024.

Textbox 1. Eligibility criteria.

Inclusion criteria

- Population: people diagnosed with major depressive disorder aged >18 years
- Concept: therapeutic guidelines for the self-management of symptoms of major depressive disorder
- Context: symptoms of major depressive disorder, such as changes in weight, appetite, and sleep; feelings of guilt; irritability or bad mood; anhedonia; fatigue; and low self-esteem
- Type of study: experimental and quasi-experimental studies (randomized and nonrandomized controlled trials), analytical observational studies (prospective and retrospective cohort studies, case-control studies, and analytical cross-sectional studies), descriptive observational studies (case series, individual case reports, and descriptive cross-sectional studies), qualitative and quantitative studies, systematic reviews, meta-analyses, scoping reviews, and literature reviews; due to the scarcity of publications found in the databases, it was also decided to include gray literature (course completion studies, dissertations, and theses)
- Year: there were no restrictions on the search period as the first search strategies developed yielded few studies on the subject
- Language: English, Spanish, and Portuguese
- Results: publications whose objective addressed therapeutic guidelines to assist in the self-management of patients with active symptoms of major depressive disorder

Exclusion criteria

- Population, concept, and context: publications without information on the population, concept, and context of interest for this study
- Type of study: duplicate studies, opinion articles, letters to the editor, abstracts of conference proceedings, and studies with unavailable text
- Language: studies in languages other than those selected
- Results: publications that did not fit the objectives and did not contain information related to the population, concept, and context of this study

Study Identification and Selection

The process of identifying and selecting studies is one of the most important stages in the conduct of a scoping review. To minimize potential bias, the *JBI Manual for Evidence Synthesis* guidelines were used and explained in a meeting with 6 researchers to ensure that this process was clear to everyone.

Thus, to maintain the rigor of the screening process, the studies obtained from each of the databases based on the eligibility criteria were exported to a reference management software (EndNote; Clarivate Analytics), and duplicates were removed. The remaining studies were then imported into the Rayyan application (Qatar Computing Research Institute) for analysis and selection based on titles and abstracts.

The titles and abstracts were read and analyzed by 2 independent reviewers to identify those that were potentially eligible. After reading the titles and abstracts, the preselected studies were subjected to data analysis and mapping, which consisted of carefully reading and classifying the texts from which the results were extracted.

The selected studies were read in full by 2 additional reviewers (MdPSdSN and PdCT) to confirm their relevance to the research question. Any doubts about the inclusion of studies would be resolved through discussion with a third reviewer until a consensus was reached. It was not necessary to contact a third reviewer.

Data Extraction and Analysis

As recommended by the JBI, the authors developed a tool to extract the data for the scoping review. Those responsible for data extraction (MdPSdSN, PdCT, and PMM) read and carefully classified the texts. The data were documented in a Microsoft Excel (Microsoft Corp) spreadsheet for better data extraction and evaluation. The final studies were determined from this step. The following variables were extracted from the studies: symptoms, authors, year, reference, study title, objective, study design, and mapping of results.

The focuses of the studies were analyzed, and the results and discussion are presented descriptively and quantified as frequencies in tables. The Discussion section synthesizes the evidence found during the review to explore it and compare it with the existing literature.

The mapped results were categorized into 7 major groups and subgroups of therapeutic orientations for the self-management of major depressive disorder based on the similarities between them. For example, therapeutic orientations focused on physical activity or regular exercise and healthy eating were classified as adoption of healthy habits, whereas listening to music, watching movies, reading, painting, going out with friends, taking outdoor walks, and practicing spiritual and religious activities were grouped in the category of leisure or pleasurable activities. This was done for all the therapeutic orientations identified considering the objective of this scoping review.

Regarding the methodological assessment (although it was not mandatory), each included study had its level of evidence identified based on the JBI appraisal tools [13] and the study design.

Thus, classification I was assigned to systematic reviews and meta-analyses of randomized clinical trials, classification II was assigned to randomized clinical trials, classification III was assigned to nonrandomized controlled trials, classification IV was assigned to case-control or cohort studies, classification V was assigned to systematic reviews of qualitative or descriptive

studies, classification VI was assigned to qualitative or descriptive studies, and classification VII was assigned to opinions of authorities or expert committee reports. This hierarchy classifies levels I and II as strong, levels III to V as moderate, and levels VI to VII as weak.

Results

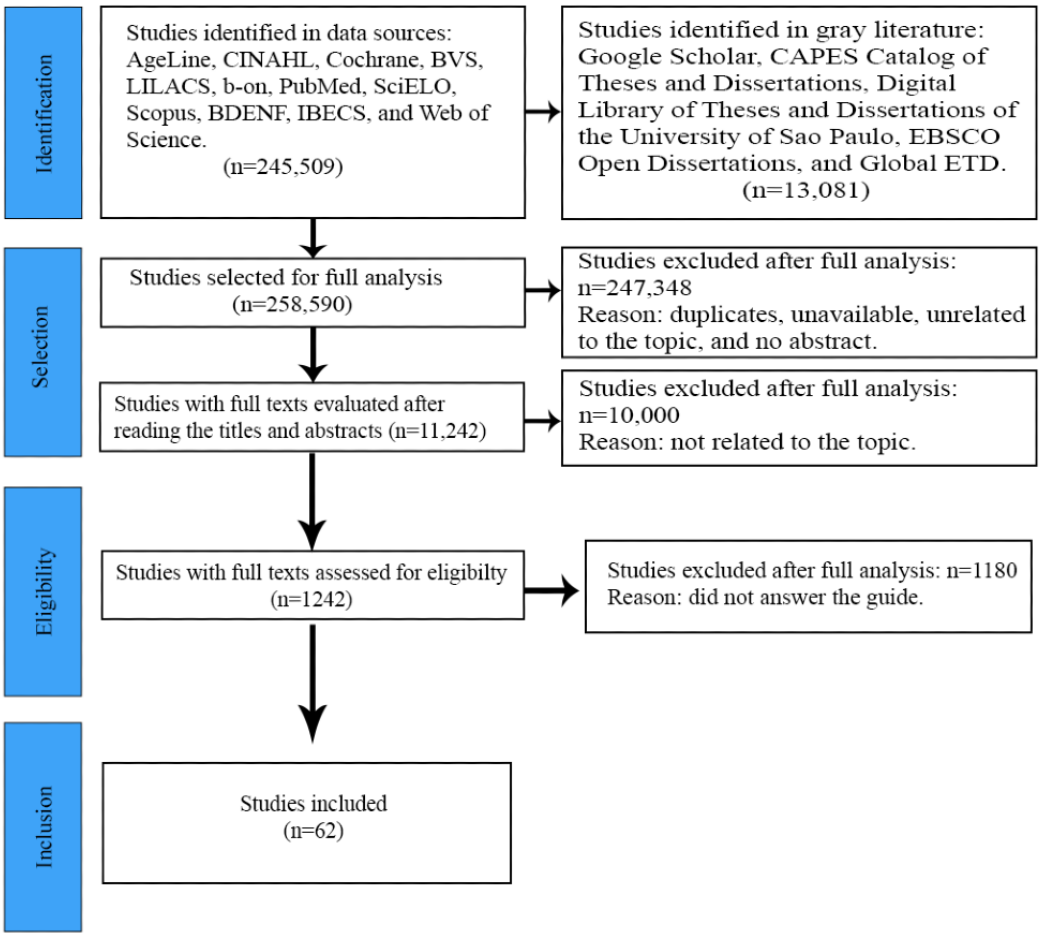
Overview of the Studies Found

Initially, 258,509 potentially eligible publications were found in the databases and 13,081 publications in the gray literature sources, of which 1538 (0.63%) focused on sleep disorders, 8315 (3.39%) focused on appetite disorders, 6732 (2.74%) focused on weight disorders, 5814 (2.37%) focused on feelings of guilt, 2509 (1.02%) focused on irritable mood, 200,002 (81.46%) focused on low self-esteem, 1848 (0.75%) focused on anhedonia, and 31,832 (7.64%) focused on fatigue.

The inclusion and exclusion criteria were applied, and 247,348 duplicate studies were removed, leaving 11,242 (4.55%). Next, a new screening was carried out in search of available publications that had abstracts and dealt with therapeutic guidelines. Of the remaining 11,242 studies, 10,000 (88.95%) were excluded, leaving 1242 (11.05%).

Once the 1242 studies had been selected, the second full reading was carried out. Of the 1242 studies, 1180 (95.01%) were excluded because they did not answer the guiding question. Thus, 62 studies were included in the final sample of this review. Figure 1 shows the results using the PRISMA-ScR flow diagram.

Figure 1. Flow diagram of the search and selection of studies according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 model.



Characteristics of the Included Studies

Of the 62 studies dealing with therapeutic guidelines (Table 1), 71% (n=44) came from indexed data sources, and 29% (n=18) were gray literature indexed on Google Scholar (n=13, 21%), doctoral theses (n=3, 5%), and Master's dissertations (n=2, 3%).

It was found that the studies were published between 2011 and 2023, with a higher number of publications in the years 2018 to 2023 (40/62, 63%), demonstrating that the use of therapeutic guidelines is recently evolving. In terms of language, most publications were in English (North America—United States) and Portuguese (South America—Brazil), totaling 63% (40/62) of the publications (Table 1).

Regarding study design, the studies were clinical trials (30/62, 48%), systematic reviews (6/62, 10%), narrative literature reviews (7/62, 11%), systematic reviews and meta-analyses (6/62, 10%), qualitative studies (3/62, 5%), pilot studies (3/62, 5%), experimental studies (2/62, 3%), case reports (2/62, 3%), noncontrolled studies (1/62, 2%), cross-sectional studies (1/62, 2%), and scoping reviews (1/62, 2%; Table 1).

Table 1. General characterization of the included studies (N=62).

Category and variable	Studies, n (%)
Year of publication	
2011-2017	22 (37)
2018-2023	40 (63)
Origin of the studies	
America	40 (63)
Europe	10 (17)
Asia	9 (15)
Oceania	2 (3)
Africa	1 (2)
Study design	
Clinical trial or experimental, pilot, qualitative, or quantitative study	38 (61)
Systematic review and meta-analysis, scoping review, or literature review	20 (32)
Noncontrolled design, case study, or cross-sectional study	4 (7)

Summary of the Extracted Variables

Table 2 presents the variables extracted from the studies: symptoms, authors, year of publication, reference, title of the study, objective, and study design. All 62 studies focused on people aged >18 years with a diagnosis of major depressive disorder and active symptoms, with a maximum age of 70 years. Of the 15 publications selected on the symptom of sleep disorders, 10 (67%) were published on PubMed [14-23], 1 (7%) was published on SciELO [24], 1 (7%) was published on AgeLine [25], and 3 (20%) were published as gray literature [26-28].

Table 2. Studies found according to symptom, study title, objective and study design.

Symptom and study	Study title	Objective	Study design
Sleep disorders; Yang [26], 2019	"Chinese medicine for major depressive disorder: clinical evidence, patients' experience and expectations, and doctors' perceptions"	To evaluate and understand the evidence on Chinese medicine for major depressive disorder in the literature	Randomized clinical trial; level II
Sleep disorders; Geoffroy et al [14], 2018	"Insomnia and hypersomnia in major depressive episode: prevalence, sociodemographic characteristics and psychiatric comorbidity in a population-based study"	Examine the frequency of sleep complaints, co-occurrences, sociodemographic characteristics, and psychiatric comorbidities associated with each type of sleep profile	Randomized clinical trial; level II
Sleep disorders; Leggett et al [15], 2018	"Bright light as a preventive intervention for depression in late-life: a pilot study on feasibility acceptability and symptom improvement"	Examine the feasibility and acceptability of a portable bright light and its impact on sleep disorders and symptoms of depressive disorders in older adults	Pilot study; level II
Sleep disorders; Luca et al [16], 2013	"Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression"	Review and discussion of sleep disorders during major depressive disorder (in particular night terrors, nightmares, hypersomnia, and insomnia)	Case report; level VI
Fatigue; Hulme et al [29], 2018	"Fatigue interventions in long term, physical health conditions: a scoping review of systematic reviews"	Map effective guidelines in all clinical conditions for the symptom of fatigue and whether these guidelines can be applied to other symptoms	Scoping review; level I
Sleep disorders; de Sousa Ibiapina et al [24], 2022	"Effects of music therapy on symptoms of anxiety and depression in adults diagnosed with mental disorders: a systematic review"	Identify and synthesize the evidence from studies that have evaluated the effects of music therapy on the symptoms of anxiety and depression in adults with mental disorders	Systematic literature review; level I
Fatigue; Macnamara et al [30], 2018	"Personalized relaxation practice to improve sleep and functioning in patients with chronic fatigue syndrome and depression: study protocol for a randomised controlled trial"	Facilitate new approaches in clinical practice and more meaningful results for patients experiencing states of chronic fatigue	Randomized clinical trial; level II
Anhedonia; Ito et al [31], 2019	"Augmentation of positive valence system-focused cognitive behavioral therapy by inaudible high-frequency sounds for anhedonia: a trial protocol for a pilot study"	Testing the effect of increasing inaudible high-frequency sounds on the effectiveness of CBT[a] focused on the positive valence system to treat anhedonia	Pilot study; level II
Anhedonia; Braun et al [32], 2019	"A pilot study investigating the effect of music-based intervention on depression and anhedonia"	To investigate the effect of a music therapy-based approach to managing major depressive disorder and its associated symptoms.	Clinical trial; level III
Anhedonia; Ebrahem and Masry [33], 2017	"Effect of relaxation therapy on depression, anxiety, stress and quality of life among diabetic patients"	To evaluate the effect of relaxation therapy on major depressive disorder and its associated symptoms such as anxiety, stress, quality of life and blood glucose levels.	Quasi-experimental study; level II
Low self-esteem; Kinser et al [34], 2013	"'A feeling of connectedness': perspectives on a gentle yoga intervention for women with major depression"	To evaluate the feasibility, acceptability, and effects of yoga among women with major depressive disorder	Randomized clinical trial; level II
Low self-esteem; Grandia [35], 2014	"Patient-initiated strategies for self-management of depression and low mood: understanding theory and changing behavior"	Examine the theory of patient-planned behavior for major depressive disorder self-management	Randomized clinical trial; level II
Sleep disorders; Prathikanti et al [25], 2017	"Treating major depression with yoga: a prospective, randomized, controlled pilot trial"	Examine a therapeutic guideline of hatha yoga for 8 weeks as monotherapy for major depressive disorder	Randomized clinical trial; level II
Guilt; Sathyanarayan et al [36], 2019	"Role of yoga and mindfulness in severe mental illnesses: a narrative review"	To review the effectiveness of yoga and mindfulness as a treatment modality for major depressive disorder	Literature review; level V
Low self-esteem; van Grieken et al [37], 2015	"Patients' perspective on self-management in the recovery from depression"	Determine therapeutic guidelines for patients to recover from major depressive disorder	Randomized clinical trial; level II
Sleep disorders; Ma et al [17], 2018	"The effects of Tai Chi on sleep quality in Chinese American patients with major depressive disorder: a pilot study"	Evaluate the effects of tai chi on sleep quality and functioning among patients with major depressive disorder	Pilot study; level II

Symptom and study	Study title	Objective	Study design
Guilt; Sarsak et al [38], 2020	"Applied occupational therapy for major depressive disorder: clinical case report"	Describe an intervention carried out via occupational therapy on an older woman with major depressive disorder and suicidal thoughts	Clinical case study; level VI
Sleep disorders; Xu et al [18], 2023	"Clinical evidence for association of acupuncture with improved major depressive disorder: a systematic review and meta-analysis of randomized control trials"	To determine the efficacy and safety of acupuncture for major depressive disorder	Meta-analysis; level I
Angry mood; Souza [39], 2019	"Chinese auricular acupuncture in the treatment of depression"	To investigate the efficacy of Chinese auricular acupuncture for major depressive disorder	Study of a qualitative nature; level V
Sleep disorders; Farrar and Farrar [19], 2020	"Clinical aromatherapy"	Investigate the history, supporting theories, guidelines, plant sources, safety, and pathophysiological responses of clinical aromatherapy in nursing	Systematic literature review; level I
Guilt; Birgitta et al [40], 2018	"Treatment of depression and/or anxiety - outcomes of a randomised controlled trial of the tree theme method® versus regular occupational therapy"	To compare therapeutic guidelines for major depressive disorder in relation to activities of daily living, psychological symptoms of major depressive disorder, and health-related aspects	Randomized clinical trial; level II
Sleep disorders; Ter Heege et al [20], 2020	"The clinical relevance of early identification and treatment of sleep disorders in mental health care: protocol of a randomized control trial"	Identify the prevalence of sleep disorders in different mental disorders, including major depressive disorder	Meta-analysis; level I
Sleep disorders; Russi et al [27], 2022	"Physical therapy intervention in the treatment of insomnia"	To evaluate a tool for restoring physiological balance and improving sleep in major depressive disorder	Experimental study; level III
Guilt; Duggal [41], 2019	"Self-management of depression: beyond the medical model"	Describe a new self-management paradigm in major depressive disorder centered on the patient beyond the improvement of clinical symptoms	Systematic literature review; level I
Weight change; Naslund et al [42], 2017	"Lifestyle interventions for weight loss among overweight and obese adults with serious mental illness: a systematic review and meta-analysis"	Evaluate the effects of a therapeutic approach to lifestyle changes on short- and long-term changes in body weight in people with major depressive disorder	Systematic review and meta-analysis; level I
Angry mood; de Cassia Rondina et al [43], 2018	"Practicing physical exercise and symptoms of depression in college students"	To investigate the relationship between major depressive disorder symptoms and physical activity patterns in university students	Randomized clinical trial; level II
Weight change; Daumit et al [44], 2013	"A behavioral weight-loss intervention in persons with serious mental illness"	Determine the effectiveness of guidelines for weight loss in adults with severe mental illness, including major depressive disorder	Essay randomized clinical trial; level II
Appetite change; Kazemi et al [45], 2020	"Effect of probiotic and prebiotic vs placebo on psychological outcomes in patients with major depressive disorder: a randomized clinical trial"	Investigate the effect of supplementation with probiotics and prebiotics on appetite, BMI, weight, and energy intake in patients with major depressive disorder	Randomized clinical trial; level II
Weight change; Colombari et al [46], 2018	"The effect to behavioral weight-loss intervention on depressive symptoms among Latino immigrants in a randomized controlled trial"	Testing the effectiveness of therapeutic guidelines for changing lifestyle life versus usual care among adults with obesity and major depressive disorder	Systematic review and meta-analysis; level I
Weight change; Carraça et al [47], 2013	"The association between physical activity and eating self-regulation in overweight and obese women"	To assess the importance of physical exercise and healthy eating for managing low self-esteem and depressed mood in people with major depressive disorder	Randomized clinical trial; level II
Sleep disorders; Rethorst et al [21], 2015	"IL-1β and BDNF are associated with improvement in hypersomnia but not insomnia following exercise in major depressive disorder"	Examine changes in hypersomnia and insomnia after increased exercise in people with major depressive disorder	Randomized clinical trial; level II
Angry mood; de Vaz Pato Oom [48], 2019	"Physical activity in the treatment of young adult depressive disease"	To evaluate the beneficial effect of physical activity in the treatment of major depressive disorder	Literature review; level V

Symptom and study	Study title	Objective	Study design
Low self-esteem; Kandola et al [49], 2019	"Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity"	Evaluate the main biological and psychosocial mechanisms of activity as an antidepressant	Systematic review; level I
Low self-esteem; Barton et al [50], 2012	"Exercise-, nature- and socially interactive-based initiatives improve mood and self-esteem in the clinical population"	Evaluate initiatives to promote health through outdoor exercise involving a population with major depressive disorder	Randomized clinical trial; level II
Weight change; Soares et al [51], 2020	"Effects of physical exercise on obesity and depression: a review"	To integrate the findings on the effects of physical exercise on obesity and major depressive disorder	Literature review; level V
Anhedonia; Brush et al [52], 2022	"A randomized trial of aerobic exercise for major depression: examining neural indicators of reward and cognitive control as predictors and treatment targets"	To demonstrate the effectiveness of aerobic exercise among adults with major depressive disorder	Randomized controlled clinical trial; level II
Anhedonia; Belvederi et al [53], 2019	"Physical exercise in major depression: reducing the mortality gap while improving clinical outcomes"	Provide a concise update on the effectiveness of physical exercise on major depressive disorder and reduction in cardiovascular mortality	Systematic literature review; level I
Anhedonia; Turner et al [54], 2019	"Physical activity and depression in MS: the mediating role of behavioral activation"	Evaluate the impact of physical activity to improve fatigue and symptoms of depressive disorders in individuals with multiple sclerosis	Randomized controlled clinical trial; level II
Anhedonia; Toups et al [55], 2017	"Exercise is an effective treatment for positive valence symptoms in major depression"	To evaluate the effect of physical exercise on the symptoms of major depressive disorder	Essay randomized clinical trial; level II
Anhedonia; Archer et al [56], 2014	"Effects of physical exercise on depressive symptoms and biomarkers in depression"	To investigate the effects of physical exercise on depressive symptoms and biomarkers in major depressive disorder	Systematic review of the literature; level I
Angry mood; Bains and Abdijadid [57], 2022	"Major depressive disorder"	Identify, describe, and review the etiology, management, and clinical presentation of major depressive disorder	Review and meta-analysis; level I
Low self-esteem; Bajaj et al [58], 2016	"Mediating role of self-esteem on the relationship between mindfulness, anxiety, and depression"	To examine the mediating effects of self-esteem on the association among mindfulness, anxiety, and major depressive disorder	Cross-sectional study; level VI
Low self-esteem; Randal et al [59], 2015	"Mindfulness and self-esteem: a systematic review"	Evaluate studies investigating the association between mindfulness and self-esteem in relation to major depressive disorder	Systematic literature review; level I
Angry mood; Farb et al [60], 2018	"Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy"	Evaluate relapse rates in patients with depression in remission receiving CBT based on mindfulness and CT[b]	Meta-analysis; level I
Sleep disorders; Kuyken et al [22], 2019	"Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials"	Conduct a meta-analysis of individual patient data to examine the effectiveness of mindfulness-based CT in the treatment of major depressive disorder	Systematic review and meta-analysis; level I
Appetite change; Katterman et al [61], 2014	"Mindfulness meditation as an intervention for binge eating, emotional eating, and weight loss: a systematic review"	Examine studies on mindfulness for the treatment of binge eating associated with major depressive disorder	Systematic review of the literature; level I
Guilt; Fawns [62], 2013	"Mindfulness based cognitive therapy and emotion focused therapy as treatments for major depression"	To describe the symptomatology and treatment options for major depressive disorder, and to examine and compare the effectiveness of two treatment orientations: mindfulness-based cognitive therapy and emotion-focused therapy	Literature review; level V
Angry mood; Economides et al [63], 2018	"Improvements in stress, affect and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial"	To evaluate the mindfulness-based Headspace smartphone app for managing symptoms of major depressive disorder	Randomized clinical trial study; level II
Appetite change; Person [64], 2021	"A two-study investigation into the link between rumination and night eating, and symptom improvement following a mindfulness-based intervention"	To explore and investigate the link between night eating syndrome and depressive symptoms	Clinical study; level II

Symptom and study	Study title	Objective	Study design
Sleep disorders; Bellingham [28], 2019	"Spiritually focused mindfulness meditation: an interpretative phenomenological analysis of the effect of spiritually focused mindfulness meditation on depression with a clinical population"	Explore the use of focused mindfulness meditation for major depressive disorder	Qualitative study; level VI
Guilt; Schanche et al [65], 2020	"The effects of mindfulness-based cognitive therapy on risk and protective factors of depressive relapse - a randomized wait-list controlled trial"	Explore the effects of mindfulness-based CT on risk and protective factors for major depressive disorder relapse in the domains of cognition, emotion, and self-relationship	Randomized clinical trial study; level II
Angry mood; Rech et al [66], 2022	"Techniques for managing the emotion of anger: a systematic review"	Identify the main techniques for managing anger in adults with major depressive disorder	Systematic review of the literature; level I
Sleep disorders; Chung et al [23], 2020	"Mobile app use for insomnia self-management in urban community-dwelling older Korean adults: retrospective intervention study"	Explore the relationship between sleep quality, memory concerns (memory loss) and depressive symptoms	Randomized clinical trial; level II
Low self-esteem; Moloud et al [67], 2022	"Cognitive-behavioral group therapy in major depressive disorder with focus on self-esteem and optimism: an interventional study"	Determine the effect of group CBT for the management of low self-esteem in patients with major depressive disorder	Randomized clinical trial study; level II
Guilt; Dobkin et al [68], 2019	"Cognitive behavioral therapy improves diverse profiles of depressive symptoms in Parkinson's disease"	Examine the impact of CBT on different depressive symptoms in Parkinson disease	Randomized clinical trial; level II
Angry mood; Santos [69], 2017	"Efficacy of procedural cognitive therapy and behavioral activation in the treatment of major depressive disorder: a randomized clinical trial"	To compare the effectiveness of psychotherapies for patients with major depressive disorder	Randomized clinical trial study; level II
Angry mood; Ahern et al [70], 2017	"Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis"	Evaluate the clinical effectiveness and evidence for the use of online CBT as an affordable treatment solution for major depressive disorder	Systematic review and meta-analysis; level I
Low self-esteem; Korrelboom et al [71], 2012	"Competitive memory training (COMET) for treating low self-esteem in patients with depressive disorders: a randomized clinical trial"	Evaluate whether competitive memory training is an effective intervention for patients with major depressive disorder	Randomized clinical trial study; level II
Anhedonia; Wang et al [72], 2020	"Guided self-help behavioral activation intervention for geriatric depression: protocol for pilot randomized controlled trial"	Pilot a therapeutic guided self-help intervention for the treatment of major depressive disorder in older adults	Randomized clinical trial; level II
Guilt; Alves and Bonvicini [73], 2022	"The role of behavioral activation in the management of depressive symptoms"	Describe and evaluate behavioral activation as a psychotherapy tool and guideline for managing mood, behavior, and emotions in patients with major depressive disorder	Systematic review; level I
Fatigue; Wisenthal et al [74], 2019	"Insights into cognitive work hardening for return-to-work following depression: qualitative findings from an intervention study"	Contribute to the literature on the effectiveness of cognitive maturation to return to work after episodes of major depressive disorder	Randomized clinical trial; level II
Weight changes; Berman et al [75], 2016	"Uncontrolled pilot study of an acceptance and commitment therapy and health at every size intervention for obese, depressed women: accept yourself!"	Evaluate the feasibility and outcomes of a new treatment based on self-acceptance for women with obesity and major depressive disorder	Randomized uncontrolled clinical trial; level II

[a]CBT: cognitive behavioral therapy.

[b]CT: cognitive therapy.

For the symptom of anhedonia, 9 publications were selected: 3 (33%) published on BVS [32,52,72], 3 (33%) published on PubMed [53-55], 2 (22%) published on Scopus [31,56], and 1 (11%) published as gray literature [33]. Similarly, the symptom of low self-esteem was identified in 9 publications: 5 (56%) published in PubMed [34,37,49,50,71], 1 (11%) published on Scopus [58], and 3 (33%) published as gray literature [35,59,67]. Furthermore, 9 publications were also selected for the symptom of irritated mood: 4 (44%) published on PubMed [57,60,63,70] and 5 (56%) published as gray literature [39,43,48,66,69].

For the symptom of guilt, 8 publications were selected: 4 (50%) published on PubMed [40,41,65,68] and 4 (50%) published as gray literature [36,38,62,73]. Regarding the symptom of weight change, 6 publications were identified: 5 (83%) published on PubMed [42,44,46,47,75] and 1 (17%) published as gray

literature [51]. For the symptom of fatigue, 3 publications were selected: 2 (67%) published on PubMed [29,30] and 1 (33%) published on CINAHL [74]. Finally, for the symptom of appetite change, 3 publications were identified: 2 (67%) published on PubMed [45,61] and 1 (33%) published as gray literature [64]. Regarding the level of evidence, 48% (30/62) of the studies were level II, 34% (21/62) were level I, 8% (5/62) were level V, 6% (4/62) were level VI, and 3% (2/62) were level III.

Self-Management of Major Depressive Disorder

The evidence available from the included studies to promote self-management of the symptoms of major depressive disorder describes the characteristics, benefits, and applicability of therapeutic guidelines in everyday life, demonstrating that these guidelines can promote a safe space to share conversations on sensitive subjects such as stigma and prejudice, adverse drug events, diagnosis, and relapse [14-75].

The findings also point out that, to promote self-management through therapeutic guidelines, it is necessary to know the functionality of the patient or family, have access to community resources and a social support network, and monitor the participation of these individuals in their care [16,33,35,37,40]. The interdisciplinary team's online or face-to-face follow-up of patients with major depressive disorder involves monitoring the use of therapeutic guidelines and whether they are being carried out correctly and appropriately considering the individual's uniqueness, as well as monitoring changes in symptoms and problems with medication, providing social support, and identifying patients at high risk of relapse [23,41].

Another common point cited in the studies was the ability of health professionals to instruct patients with major depressive disorder on how to deal with symptoms and biosociopsychological demands [16,31-33]. Professionals require skills and good communication with patients when instructing them on the use of therapeutic guidelines for managing the disorder, since incorrect instructions or misunderstandings may make the application of therapeutic guidelines less efficient [23,35,37,40,67,70,74,75]. If the professional advises the patient to sunbathe (phototherapy) during the day, but does not specify the time, this may cause the patient to take this guidance at inappropriate times and be harmful to their health. Therefore, phototherapy would not be effective. The findings encourage a multidisciplinary approach to the applicability of therapeutic guidelines, focusing on the specificities of each person and collective actions, strengthening the patient's ability to self-manage.

Therapeutic Guidelines for Self-Management

The analysis of the studies, as described previously, resulted in the mapping and categorization of 7 large groups of therapeutic guidelines for the self-management of major depressive disorder. These guidelines are made up of a set of actions or activities (n=40), which have been called subinterventions based on the similarities between them (Table 3).

In total, 52% (32/62) of the therapeutic guidelines identified fell into the category of psychotherapy, which helps monitor and plan activities with the person (in the context of their symptoms), manage adverse experiences, and develop social skills [14,16,17,22,23,26,28,29,31,36-38,46,56,58-72,74,75]. Adoption of healthy habits (25/62, 40%) and integrative and complementary practices (17/62, 27%) were the second and third most observed categories. The process of adopting integrative and complementary practices [17,19,20,24-26,29-40], together with healthy habits such as eating a balanced diet and exercising frequently and for the right duration, showed a good response in improving depressive conditions, as well as being an important factor in preventing relapses of major depressive disorder [21,26,27,29,30,33,37,38,41-57].

Finally, relaxation techniques (9/62, 14%), consultation with a health professional (14/62, 22%), pharmacological therapy (9/62, 14%), and leisure or pleasurable activities (4/62, 6%) were the least observed therapeutic guidelines in the studies. However, they offered promising insights into self-management of the disorder and underscored the importance of a multifaceted approach to managing this complex condition [16,20,23,26-30,33,36-38,41,50,57,66].

It is widely recognized that the concept of self-management is traditionally associated with approaches that the patient can carry out autonomously, such as adopting healthy habits, physical exercise, and relaxation techniques. However, it is crucial to emphasize that, in the context of major depressive disorder, self-management transcends these independent symptom management practices [37].

It involves empowering the patient to take active control and gain an in-depth understanding of their condition through the application of a variety of therapeutic interventions [35]. Patients with major depressive disorder may experience feelings of frustration and a range of negative thoughts, see everyday problems as major catastrophes, have difficulty recognizing efforts that awaken hope for life, and cope with the possible adverse events that psychiatric medications can cause [66].

For this reason, orientation toward psychotherapy and psychopharmaceuticals is justified as a self-management strategy as a way of coping with the adverse effects of medication and all the feelings of self-demand, excessive demands, and frustration caused by the disorder.

In relation to the population to which this study refers, the results show that patients with major depressive disorder have motivational, cognitive, and psychological deficits that can modify their ability to cope and reason [22,60,63,65].

In view of this, self-management of the signs and symptoms resulting from major depressive disorder needs to take place at the beginning of treatment, with the support of family or caregivers; friends; community organizations; and, most especially, the multi-professional health team to help with the correct use of therapeutic options until the person is able to self-manage independently [14-75]. Therefore, the therapeutic guidelines found in this review are not restricted to the professional-patient relationship; these relationships are broad and include families and the community.

Table 3. Mapping of the results regarding the therapeutic guidelines.

Therapeutic guideline	Subinterventions
Adopting healthy habits	• Physical activity or regular exercise [21,26,27,29,30,33,37,38,41-57] • Healthy eating [26,27,29,37,41-46]
Leisure or pleasurable activities	• Listening to music, watching movies, reading, painting, going out with friends, taking outdoor walks, and practicing spiritual and religious activities [37,38,41,50]
Relaxation techniques	• Progressive muscle relaxation [33,41,66] • Respiratory rehabilitation [16,26,27,29,30,36] • Stretching [16,26,27,29,41,66] • Warm baths [16] • Massages [29,41,66] • Meditation [30]
Pharmacological therapy	• Psychopharmacological therapy [16,20,26,28,29,36,38,41,57]
Consultation with a health professional	• Psychoeducation or professional counseling [16,20,23,26,27,29,36-38,41-43,57,66]
Psychotherapy	• Behavioral activation [69,73] • Cognitive behavioral psychotherapy [14,16,23,26,29,31,38,58,67-71] • Mindfulness [22,26,28,36-38,59-66] • Psychotherapy focused on problem-solving [29,38,74] • Competitive memory training psychotherapy [71] • Self-efficacy psychotherapy [46] • Guided self-help psychotherapy [23,72] • Interpersonal psychotherapy [17,23,56] • Therapy focused on acceptance, commitment [62], and emotion [75]
Integrative and complementary practices	• Acupuncture [19,26,29,39] • Music therapy [24,30-33] • Phytotherapy [26,29] • Phototherapy [26,29] • Aromatherapy [19] • Cryotherapy [29] • Thermotherapy [29] • Tai chi [17,29,38] • Reflexology [29] • Qigong [29] • Yoga [25,29,34-37] • Art therapy [40] • Chromotherapy [20]

Discussion

Principal Findings

From this review, it was possible to identify the therapeutic guidelines for the self-management of the symptoms of major depressive disorder in scientific production as psychotherapy, adoption of healthy habits, integrative and complementary practices, relaxation techniques, consultation with a health professional, pharmacological therapy, and leisure or pleasurable activities.

The selected studies show that knowledge on the subject significantly improves self-management of the symptoms of major depressive disorder through these therapeutic guidelines to promote behavior changes and awareness of the symptoms of the disorder, prevention of relapses, reduction of the perception of obstacles, and increased adherence to treatment.

A small number of articles published in journals on the subject were found for all the symptoms investigated, especially irritable mood, followed by fatigue, weight change, appetite change, sleep changes, and low self-esteem, necessitating the addition of secondary documents that considerably broadened the identification of other therapeutic guidelines.

Most of the studies originated in America (40/62, 63%), where the largest number of people with major depressive disorder is concentrated. Regarding the concept that triggered the research, the studies involved health professionals, students, and the general population, covering the definition, description, comparative analysis, efficiency, and applicability of therapeutic guidelines. These studies were included to increase the possibility of nurses playing a leading role in the self-management of symptoms of major depressive disorder [14-75].

The factors considered important for the self-application of therapeutic guidelines in the studies were cognition; level of education; environmental factors; and the functional capacity of the patient, caregivers, and family members. It should be emphasized that, for self-management, clients who are going

to follow these therapeutic guidelines should receive them in clear and objective language and format, with specificity and relevance for each symptom, enabling adherence to treatment and favoring the therapist–patient or caregiver bond [26]. These therapeutic guidelines are also considered low cost and easy to use, optimize rehabilitation and autonomy, and encourage self-care [26]. It is important to highlight the definitions of each of the therapeutic guidelines identified in this review.

Integrative and complementary practices, as therapeutic resources based on traditional knowledge, were the therapeutic guidelines most commonly found in the literature as effective for the self-management and management of major depressive disorder symptoms, and their applicability aimed to induce a state of harmony and balance throughout the body.

Phototherapy has been shown to reduce depressive symptoms and improve irregularities in sleep patterns and quality, especially in older adults with major depressive disorder, as long as it is used with a certain frequency and is administered before 10 AM [14-16,26]. It has also been used in the control of nocturnal hyperphagia and has shown reductions in the percentage of food eaten after dinner and in the number of nocturnal meals eaten per week, reducing the rate of overweight and obesity, as well as being used in the control of fatigue, where it has shown a reduction in the feeling of tiredness [29].

Music therapy applied to adults [24] has been shown to be effective in physical and mental relaxation, reducing anxious and depressive symptoms and promoting well-being in a conscious and healthy way [30]. Listening to soothing noises and classical music associated with rhythmic breathing produced a significantly higher rate of adherence to major depressive disorder treatment, especially in the patients with the highest severity [31-33].

A study has shown that anhedonia and clinically significant depressive symptoms can be resistant to standard treatment but adjuvant treatment with high-frequency inaudible sound therapy increases the reward of related brain circuits and has a synergistic effect on anhedonia [31]. Integrating music therapy with conventional major depressive disorder treatment (therapies and medication) gives people the opportunity to get in touch with their emotions and provides distraction and a means of communication capable of overcoming barriers and limits to verbal expression [30].

Yoga [29,34,35] has been a beneficial exercise for managing sleep quality and other symptoms of major depressive disorder [25,36]. It is an important tool for promoting self-care and, consequently, care for everything around the patient [37]. A study carried out to explore the experiences of patients with major depressive disorder showed that practicing yoga as a self-management strategy helped them gain better insights into their own condition, improving the quality of their health care [37].

Tai chi has been shown to be more effective than traditional rehabilitation (psychotropic drugs and therapies) in relation to insomnia, with an improvement in the quality, duration, and efficiency of sleep and a reduction in thoughts of guilt associated with major depressive disorder [17,29,38]. Acupuncture has been shown to be effective in treating people with major depressive disorder [18,39], mainly in reducing sleep disorders but also in symptoms related to changes in appetite, irritability, anxiety, and sadness, and has not been shown to be effective for suicidal ideation [26,29].

With regard to herbal medicines, St John's wort, *chai hu*, and *gancao* were the most widely used. In 3% (2/62) of the studies, after 2 weeks of administering these herbs (alone or in combination) as an alternative to conventional antidepressants, they produced a statistically significant improvement in symptoms of altered sleep [26,29]. Although the effects of these herbs are not fully understood, it is likely that they produce antidepressant effects through multiple pathways or targets that interact with each other.

To a lesser extent among the studies but no less importantly, aromatherapy [19], cryotherapy, thermotherapy, reflexology, qigong [29], art therapy [40], and chromotherapy [20] improved empowerment in the search for self-care and taking responsibility for one's own health, as well as reducing levels of anxiety and stress, improving sleep disorders, improving the immune system, and lowering blood pressure levels [29].

People with major depressive disorder have shown improvements in sleep disorders, anxiety, and stress, as well as enhancements in immune system function and reductions in blood pressure levels when using respiratory rehabilitation [16,26,27,29,30,36], stretching [16,26,27,29,41,66], warm baths [16], massages [29,41,66], acupuncture, herbal medicine, phototherapy, reflexology, qigong, and yoga [17,19,20,24,26,29,34,38-40] as therapeutic guideline.

Cryotherapy and thermotherapy are not commonly used to treat the symptoms of major depressive disorder, but they help relax the body and mind as a physiotherapeutic resource [16,29]. There was a significant improvement in stress and body pain related to fatigue, anxiety, and swelling in the feet and legs [29].

Another important therapeutic guideline was the adoption of healthy lifestyle habits, such as healthy eating [17,26,27,29,41-46] and physical exercise [21,26,27,30,33,37,38,41-57]. These reduce stress; improve mood, body image perception, and self-esteem; stimulate cognitive functioning; promote greater satisfaction with life [45,46]; and can help people feel stronger and more capable, helping reduce the feelings of hopelessness and helplessness associated with major depressive disorder [53,54].

It should be noted that, regardless of the physical activity chosen, people with major depressive disorder need to make it part of their routine and do it [47] according to their tolerance and state of health [54]. An important concern in relation to physical activity for people with major depressive disorder is the fact that some of the common symptoms of depression (fatigue, lack of energy, psychomotor retardation, despair, and feelings of worthlessness) interfere with the motivation to exercise [57].

In more severe cases, practicing physical activity can be difficult for people with major depressive disorder [55]. This can compromise adherence and long-term permanence in the exercise program [52]. It is recommended to prescribe structured

exercises based on the activities already practiced by the patient to reduce obstacles [33].

Another group of therapeutic guidelines that were also commonly identified in this review were psychotherapies. Mindfulness-based therapy [58,59] is a promising therapeutic approach for treating major depressive disorder and is part of the list of integrative and complementary practices [26,37,60]. It combines elements of therapeutic guidelines [22,61] with the aim of helping people develop a full awareness of the present moment, accepting and acknowledging their thoughts, emotions, and bodily sensations without judgment [62] and learning to observe their negative thoughts and dysfunctional thought patterns without allowing themselves to be involved with or controlled by them [63].

Mindfulness improves self-management of prodromal symptoms [64], represents a stability factor when practiced frequently and in a supervised manner [28], positively influences memory and distorted feelings of guilt, increases cognitive resilience, balances mood and sleep, reduces high levels of stress, and improves self-esteem and anxiety [65] associated with major depressive disorder relapse.

Similarly, cognitive behavioral psychotherapy [26,66] uses positive reinforcement techniques and systematic desensitization, helping people cope with stress, face the challenges of everyday life, change negative thought patterns and behaviors to more realistic and positive ones [23,67], and identify the first signs of relapse and prevent it as it focuses on social skill training [68,69]. Recently, online cognitive behavioral psychotherapy has been shown to be effective for depressed and angry mood as a cost-effective treatment modality for major depressive disorder [70].

Other psychotherapies aimed at influencing the person with major depressive disorder and helping them modify emotional, cognitive, and behavioral problems were self-efficacy psychotherapy [46]; competitive memory training [71]; guided self-help psychotherapy [23,72]; behavioral activation [69,73]; interpersonal psychotherapy [38,57,67]; psychotherapy focused on problem-solving [29,38,74]; and therapy focused on acceptance, commitment, and emotion [62,75].

Leisure time or pleasurable activities as a social practice, contrary to what is often thought, are not only carried out during the summer or on vacation but also in between daily obligations [37]. Thus, they can take the form of individual or group dynamics, whether it is reading a book, listening to music, or going for a walk with friends [37,50]. The relationship between leisure and health has fostered new techniques for the treatment of depressive disorders with the aim of enabling the individual's psychic and social readaptation [37,50].

Painting for leisure, for example, can promote the manifestation of feelings through emotions expressed verbally or not, thus helping understand the mind and its sorrows [38]. In this sense, of the possibilities for leisure activities recommended by the professionals who collaborated with the data found, art therapy [40], group and individual outings [41,50], trips [38], reading books and watching movies [37,38], listening to music [31], and attending religious institutions [50] were the most cited in the studies.

In the context of pharmacological treatment, antidepressants associated with psychotherapies [26,29,37] should be administered with caution and under medical supervision [20,38,41] based on the therapeutic alliance, clinical history, monitoring and reassessment of psychiatric conditions, and adequacy of the diagnosis, and guidance should be provided for families [28,57].

In 5% (3/62) of the studies, the association of psychotropic drugs with therapeutic guidelines was highlighted (due to the severity of the major depressive disorder clinical picture) even in the face of difficulties with medication adherence for different reasons [16,29,38], which should be anticipated and addressed proactively in every contact with the patient. Approximately 30% to 40% of people using antidepressants still do not respond as expected to treatment [27,41].

Relaxation techniques appeared in this review as a more appropriate way of coping with the psychological, environmental, and social stimuli that people with major depressive disorder may face. Among these strategies, breathing techniques [16,26,27,29,30,36], progressive muscle relaxation [33,41,66], massages [29,41,66], stretching [16,26,27,29,66], a warm bath [16], and meditation [30] were the most cited.

In the context of major depressive disorder, these techniques promote muscle relaxation, relieve stress and tension, rebalance emotions, produce hormones such as endorphins, provide a sense of well-being, and significantly improve sleep quality [16]. Incorporating these practices throughout the day or even before going to bed helps people let go of tiredness, become aware of the present moment, and pay attention to their own bodies [16,66].

Finally, therapeutic guidelines for psychoeducation or professional counseling can be used to help with self-management as major depressive disorder generates changes in the family system and structure [16,18,26,36,38], resulting in the need for clarification to relieve anxieties and doubts to improve the psychological well-being of the patient [20,37,42,43].

Thus, the wide variety of therapeutic guidelines found can increase knowledge about the symptoms and aspects of major depressive disorder, allowing the person to self-manage symptoms such as fatigue, sleep disturbances, anhedonia, and mood swings [16,27,29,66] in a more responsible and autonomous way.

These therapeutic guidelines are characterized by providing information on the diagnosis, etiology, prognosis, and course of the illness; the identification of early signs of crisis; the importance of adherence to medication or psychosocial treatment; the promotion of healthy habits and regularity in lifestyle (sleep, diet, physical activity, and substance use); and how to deal with the stigmatization, doubts, fears, and myths regarding major depressive disorder [23,41,42].

Therefore, the implications for practice are based on the fact that therapeutic guidelines help in the self-management of major

depressive disorder through behavioral, social, and emotional changes that allow for better construction of clinical reasoning, adaptation, autonomy, coping, and improvement of general health.

Limitations

The limitations of this scoping review are related to the nature of the review itself as its aim was to provide an overview of therapeutic guidelines, which may not be sufficient for the self-management of major depressive disorder as it is a complex clinical condition with a heterogeneity of symptoms. Furthermore, despite efforts to develop a comprehensive search strategy, it was difficult to find controlled and uncontrolled descriptors for the term "therapeutic guidelines," which is not an indexed descriptor.

Conclusions

By analyzing the methodological approach adopted in this study, and in accordance with the proposed objectives, it was possible to highlight a set of therapeutic guidelines to support a person's self-management in the context of major depressive disorder symptoms. This does not preclude further study of their effectiveness in the self-management of major depressive disorder symptoms but demonstrates their contribution to promoting self-management of major depressive disorder and their power to prevent possible complications.

Self-management is a promising strategy that emphasizes the person's responsibility in the care process. It goes beyond participation in health care interactions and includes dealing with symptoms and disability; managing medication and monitoring indicators; maintaining adequate levels of nutrition and exercise; and adjusting to psychological, social, and lifestyle demands.

These therapeutic guidelines can be applied by all members of a multi-professional mental health team, including nurses, provided they are trained as resources to help the person acquire new knowledge, behaviors, and skills for self-management. These guidelines are also an important resource for health professionals in their role as facilitators of a person's development of knowledge and skills to ensure their capacity for self-management.

Acknowledgments

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel, Brazil (finance code 001) and by National Council for Scientific and Technological Development Ministry of Science, Technology and Innovation call 10/2023, Track A—Emerging Groups (process 403740/2023-1). The funders had no role in the study design; data collection, analysis, and interpretation; writing of the manuscript; or decision to publish. Generative artificial intelligence was not used in the writing of this manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Authors' Contributions

PdCT contributed to the conceptualization, data curation, investigation, methodology, resource management, validation, and software implementation. PdCT was also responsible for visualization and contributed to both drafting the original manuscript and reviewing and editing it; PMM was responsible for data curation, formal analysis, and investigation; DMP contributed to investigation, methodology, and resource management; MAMJ was responsible for validation and visualization; MAdL contributed to data curation and formal analysis. CSF was involved in data curation, formal analysis, supervision, and drafting the original manuscript; CSVdBS contributed to data curation, formal analysis, and drafting the original manuscript. MdPSdSN contributed to the conceptualization of the study, data curation, funding acquisition, investigation, methodology, project administration, resource management, supervision, validation, and visualization. MdPSdSN also played a key role in drafting the original manuscript and in reviewing and editing it. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 85 KB - ijmr_v14i1e63959_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 40 KB - ijmr_v14i1e63959_app2.docx]

References

1. Mônego BG, Fonseca RP, Teixeira AL, Barbosa IG, Souza LC, Bandeira DR. Major depressive disorder: a comparative study on social-emotional cognition and executive functions. Psic Teor e Pesq 2022;38(3):27. [doi: 10.1590/0102.3772e38217.en]
2. Sousa NF, Barros MB, Medina LP, Malta DC, Szwarcwald CL. Association of major depressive disorder with chronic diseases and multimorbidity in Brazilian adults, stratified by gender: 2019 National Health Survey. Rev Bras Epidemiol 2021;24(suppl 2):e210015 [FREE Full text] [doi: 10.1590/1980-549720210015.supl.2] [Medline: 34910069]
3. GHDx: global health data exchange results tool. Institute for Health Metrics and Evaluation. 2019. URL: https://ghdx.healthdata.org/ [accessed 2023-04-30]
4. Bernhardt M, Klauke S, Schröder A. Longitudinal course of cognitive function across treatment in patients with MDD: a meta-analysis. J Affect Disord 2019 Apr 15;249:52-62. [doi: 10.1016/j.jad.2019.02.021] [Medline: 30753954]
5. Saraceno B. Citizenship as a form of tolerance. Rev Ter Ocup Univ São Paulo 2011 Aug 01;22(2):3-101. [doi: 10.11606/issn.2238-6149.v22i2p93-101]
6. Dutra VF, Bossato HR, Oliveira RM. Mediating autonomy: an essential care practice in mental health. Esc Anna Nery 2017;21(3):33. [doi: 10.1590/2177-9465-ean-2016-0284]
7. Bauer M, Whybrow PC, Angst J, Versiani M, Möller HJ. World federation of societies of biological psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, part 2: maintenance treatment of major depressive disorder and treatment of chronic depressive disorders and subthreshold depressions. World J Biol Psychiatry 2002 Apr 12;3(2):69-86. [doi: 10.3109/15622970209150605] [Medline: 12479080]
8. Nóbrega MD, Tibúrcio PC, Fernandes MC, Fernandes CS, Santos CS, Magalhães BM. Exploring the use of mobile applications for self-management of mental health treatment: scoping review. Rev Enferm UFSM 2021 Jul 22;11:e56. [doi: 10.5902/2179769264393]
9. do Carmo Barbosa AL, de São Joaquim BS. The practice of therapeutic guidance in the daily life of the rehabilitation professional. Escola de Educação Física, Fisioterapia e Terapia Ocupacional. URL: http://www.eeffto.ufmg.br/eeffto/biblioteca/monografia/560/ [accessed 2024-04-29]
10. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; Jul 27, 2016:513-551.
11. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
12. Orientações terapêuticas para autogestão do Transtorno Depressivo Maior: revisão de escopo. Open Science Framework. URL: https://osf.io/z3ya7/ [accessed 2024-04-29]
13. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2024.
14. Geoffroy PA, Hoertel N, Etain B, Bellivier F, Delorme R, Limosin F, et al. Insomnia and hypersomnia in major depressive episode: prevalence, sociodemographic characteristics and psychiatric comorbidity in a population-based study. J Affect Disord 2018 Jan 15;226:132-141. [doi: 10.1016/j.jad.2017.09.032] [Medline: 28972930]
15. Leggett AN, Conroy DA, Blow FC, Kales HC. Bright light as a preventive intervention for depression in late-life: a pilot study on feasibility, acceptability, and symptom improvement. Am J Geriatr Psychiatry 2018 May;26(5):598-602 [FREE Full text] [doi: 10.1016/j.jagp.2017.11.007] [Medline: 29269208]
16. Luca A, Luca M, Calandra C. Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression. Clin Interv Aging 2013;8:1033-1039 [FREE Full text] [doi: 10.2147/CIA.S47230] [Medline: 24019746]
17. Ma Y, Yeung A, Yang AC, Peng C, Clain A, Alpert J, et al. The effects of Tai Chi on sleep quality in Chinese American patients with major depressive disorder: a pilot study. Behav Sleep Med 2018 Sep 27;16(4):398-411. [doi: 10.1080/15402002.2016.1228643] [Medline: 27676270]
18. Xu G, Xiao Q, Huang B, Lei H, Yin Z, Huang L, et al. Clinical evidence for association of acupuncture with improved major depressive disorder: a systematic review and meta-analysis of randomized control trials. Neuropsychobiology 2023 Oct 21;82(1):1-13. [doi: 10.1159/000527903] [Medline: 36549277]
19. Farrar AJ, Farrar FC. Clinical aromatherapy. Nurs Clin North Am 2020 Dec;55(4):489-504 [FREE Full text] [doi: 10.1016/j.cnur.2020.06.015] [Medline: 33131627]
20. Ter Heege FM, Mijnster T, van Veen MM, Pijnenborg GH, de Jong PJ, Boersma GJ, et al. The clinical relevance of early identification and treatment of sleep disorders in mental health care: protocol of a randomized control trial. BMC Psychiatry 2020 Jun 24;20(1):331 [FREE Full text] [doi: 10.1186/s12888-020-02737-3] [Medline: 32580724]
21. Rethorst CD, Greer TL, Toups MS, Bernstein I, Carmody TJ, Trivedi MH. IL-1β and BDNF are associated with improvement in hypersomnia but not insomnia following exercise in major depressive disorder. Transl Psychiatry 2015 Aug 04;5(8):e611 [FREE Full text] [doi: 10.1038/tp.2015.104] [Medline: 26241349]
22. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA Psychiatry 2016 Jun 01;73(6):565-574 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]

23. Chung K, Kim S, Lee E, Park JY. Mobile app use for insomnia self-management in urban community-dwelling older Korean adults: retrospective intervention study. JMIR Mhealth Uhealth 2020 Aug 24;8(8):e17755 [FREE Full text] [doi: 10.2196/17755] [Medline: 32831177]
24. de Sousa Ibiapina AR, Lopes-Junior LC, Veloso LU, Costa AP, da Silva Júnior FJ, e Silva Sales JC, et al. Effects of music therapy on anxiety and depression symptoms in adults diagnosed with mental disorders: a systematic review. Acta Paul Enferm 2022;35:e2212. [doi: 10.37689/acta-ape/2022AR0]
25. Prathikanti S, Rivera R, Cochran A, Tungol JG, Fayazmanesh N, Weinmann E. Treating major depression with yoga: a prospective, randomized, controlled pilot trial. PLoS One 2017 Mar 16;12(3):e0173869 [FREE Full text] [doi: 10.1371/journal.pone.0173869] [Medline: 28301561]
26. Yang L. Chinese medicine for major depressive disorder: clinical evidence, patients' experience and expectations, and doctors' perceptions. Beijing University of Chinese Medicine. 2019. URL: https://tinyurl.com/39hrasy4 [accessed 2024-07-29]
27. Russi MA, Sperandio RD, De Oliveira CR, Sanches LR. Physical therapy intervention in the treatment of insomnia. Braz J Dev 2022 Nov 03;8(11):70997-71015. [doi: 10.34117/bjdv8n11-017]
28. Bellingham GL. Spiritually focused mindfulness meditation: an interpretative phenomenological analysis of the effect of spiritually focused mindfulness meditation on depression with a clinical population. Liverty University. 2019. URL: https://tinyurl.com/8hvh29n9 [accessed 2023-04-09]
29. Hulme K, Safari R, Thomas S, Mercer T, White C, Van der Linden M, et al. Fatigue interventions in long term, physical health conditions: a scoping review of systematic reviews. PLoS One 2018 Oct 12;13(10):e0203367 [FREE Full text] [doi: 10.1371/journal.pone.0203367] [Medline: 30312325]
30. Macnamara CL, Cvejic E, Parker GB, Lloyd AR, Lee G, Beilharz JE, et al. Personalised relaxation practice to improve sleep and functioning in patients with chronic fatigue syndrome and depression: study protocol for a randomised controlled trial. Trials 2018 Jul 11;19(1):371 [FREE Full text] [doi: 10.1186/s13063-018-2763-8] [Medline: 29996933]
31. Ito M, Miyamae M, Yokoyama C, Yamashita Y, Ueno O, Maruo K, et al. Augmentation of positive valence system-focused cognitive behavioral therapy by inaudible high-frequency sounds for anhedonia: a trial protocol for a pilot study. JAMA Netw Open 2019 Nov 01;2(11):e1915819 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15819] [Medline: 31747035]
32. Braun JT, Al Shirawi MI, Rotzinger S, Kennedy SH, Bartel L. A pilot study investigating the effect of music-based intervention on depression and anhedonia. Front Psychol 2019 May 8;10:1038 [FREE Full text] [doi: 10.3389/fpsyg.2019.01038] [Medline: 31133945]
33. Ebrahem SM, Masry SE. Effect of relaxation therapy on depression, anxiety, stress and quality of life among diabetic patients. Clin Nurs Stud 2017 Jan 19;5(1):35. [doi: 10.5430/cns.v5n1p35]
34. Kinser PA, Bourguignon C, Taylor AG, Steeves R. "A feeling of connectedness": perspectives on a gentle yoga intervention for women with major depression. Issues Ment Health Nurs 2013 Jun 28;34(6):402-411 [FREE Full text] [doi: 10.3109/01612840.2012.762959] [Medline: 23805925]
35. Grandia PM. Patient-initiated strategies for self-management of depression and low mood: understanding theory and changing behaviour. Thesis University of Ottawa. 2015. URL: https://ruor.uottawa.ca/items/f5a6ccef-9bfd-4eee-bddb-d3418bda5e9a [accessed 2023-02-22]
36. Sathyanarayanan G, Vengadavaradan A, Bharadwaj B. Role of yoga and mindfulness in severe mental illnesses: a narrative review. Int J Yoga 2019;12(1):3-28 [FREE Full text] [doi: 10.4103/ijoy.IJOY_65_17] [Medline: 30692780]
37. van Grieken RA, Kirkenier AC, Koeter MW, Nabitz UW, Schene AH. Patients' perspective on self-management in the recovery from depression. Health Expect 2015 Oct 02;18(5):1339-1348 [FREE Full text] [doi: 10.1111/hex.12112] [Medline: 23910797]
38. Sarsak HI. Applied occupational therapy for major depressive disorder: clinical case report. MOJ Clin Med Case Rep 2020;10(2):31-34. [doi: 10.15406/mojcr.2020.10.00337]
39. Souza CD. Chinese auricular acupuncture in the treatment of depression. University of São Paulo. 2019. URL: https://www.teses.usp.br/teses/disponiveis/22/22131/tde-31012020-111959/pt-br.php [accessed 2024-04-30]
40. Birgitta GA, Wagman P, Hedin K, Håkansson C. Treatment of depression and/or anxiety - outcomes of a randomised controlled trial of the tree theme method® versus regular occupational therapy. BMC Psychol 2018 May 23;6(1):25 [FREE Full text] [doi: 10.1186/s40359-018-0237-0] [Medline: 29792226]
41. Duggal HS. Self-management of depression: beyond the medical model. Perm J 2019;23:55 [FREE Full text] [doi: 10.7812/TPP/18-295] [Medline: 31167709]
42. Naslund JA, Whiteman KL, McHugo GJ, Aschbrenner KA, Marsch LA, Bartels SJ. Lifestyle interventions for weight loss among overweight and obese adults with serious mental illness: a systematic review and meta-analysis. Gen Hosp Psychiatry 2017 Jul;47:83-102 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.04.003] [Medline: 28807143]
43. de Cassia Rondina R, Aragão Martins R, da Silva AL, Mascarenhas GP, Travenssolo JG, Luiz dos Santos B. Practicing physical exercise and symptoms of depression in college students. Mundo Saúde 2018 Jul 30;42(3):710-727 [FREE Full text] [doi: 10.15343/0104-7809.20184203710727]

44. Daumit GL, Dickerson FB, Wang NY, Dalcin A, Jerome GJ, Anderson CA, et al. A behavioral weight-loss intervention in persons with serious mental illness. N Engl J Med 2013 Apr 25;368(17):1594-1602 [FREE Full text] [doi: 10.1056/NEJMoa1214530] [Medline: 23517118]
45. Kazemi A, Noorbala AA, Azam K, Eskandari MH, Djafarian K. Effect of probiotic and prebiotic vs placebo on psychological outcomes in patients with major depressive disorder: a randomized clinical trial. Clin Nutr 2019 Apr;38(2):522-528. [doi: 10.1016/j.clnu.2018.04.010] [Medline: 29731182]
46. Colombari FS, Stafford RS, Heaney CA, Rosas LG. The effect of a behavioral weight-loss intervention on depressive symptoms among Latino immigrants in a randomized controlled trial. J Immigr Minor Health 2018 Oct 16;20(5):1182-1189 [FREE Full text] [doi: 10.1007/s10903-017-0663-8] [Medline: 29038966]
47. Carraça EV, Silva MN, Coutinho SR, Vieira PN, Minderico CS, Sardinha LB, et al. The association between physical activity and eating self-regulation in overweight and obese women. Obes Facts 2013;6(6):493-506 [FREE Full text] [doi: 10.1159/000356449] [Medline: 24217426]
48. de Vaz Pato Oom MC. Physical activity in the treatment of young adult depressive disease. University of Beira Interior. URL: http://hdl.handle.net/10400.6/10748 [accessed 2023-04-14]
49. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. Neurosci Biobehav Rev 2019 Dec;107:525-539. [doi: 10.1016/j.neubiorev.2019.09.040] [Medline: 31586447]
50. Barton J, Griffin M, Pretty J. Exercise-, nature- and socially interactive-based initiatives improve mood and self-esteem in the clinical population. Perspect Public Health 2012 Mar 07;132(2):89-96. [doi: 10.1177/1757913910393862] [Medline: 22616429]
51. Soares TD, Peroza LR, Cerezer M, Nedel SS, Branco JC. Effects of physical exercise on obesity and depression: a review. Braz J Obes Nutr Weigh Loss 2020;14(86):511-518 [FREE Full text] [Medline: 1981-9919]
52. Brush CJ, Hajcak G, Bocchine AJ, Ude AA, Muniz KM, Foti D, et al. A randomized trial of aerobic exercise for major depression: examining neural indicators of reward and cognitive control as predictors and treatment targets. Psychol Med 2022 Apr;52(5):893-903. [doi: 10.1017/S0033291720002573] [Medline: 32838817]
53. Belvederi MM, Ekkekakis P, Magagnoli M, Zampogna D, Cattedra S, Capobianco L, et al. Physical exercise in major depression: reducing the mortality gap while improving clinical outcomes. Front Psychiatry 2018 Jan 10;9:762 [FREE Full text] [doi: 10.3389/fpsyt.2018.00762] [Medline: 30687141]
54. Turner AP, Hartoonian N, Hughes AJ, Arewasikporn A, Alschuler KN, Sloan AP, et al. Physical activity and depression in MS: the mediating role of behavioral activation. Disabil Health J 2019 Oct;12(4):635-640. [doi: 10.1016/j.dhjo.2019.04.004] [Medline: 31097413]
55. Toups M, Carmody T, Greer T, Rethorst C, Grannemann B, Trivedi MH. Exercise is an effective treatment for positive valence symptoms in major depression. J Affect Disord 2017 Feb;209:188-194 [FREE Full text] [doi: 10.1016/j.jad.2016.08.058] [Medline: 27936452]
56. Archer T, Josefsson T, Lindwall M. Effects of physical exercise on depressive symptoms and biomarkers in depression. CNS Neurol Disord Drug Targets 2014 Jan 22;13(10):1640-1653. [doi: 10.2174/1871527313666141130203245] [Medline: 25470398]
57. Bains N, Abdijadid S. Major depressive disorder. StatPearls. URL: https://www.ncbi.nlm.nih.gov/books/NBK559078/ [accessed 2024-06-23]
58. Bajaj B, Robins RW, Pande N. Mediating role of self-esteem on the relationship between mindfulness, anxiety, and depression. Pers Individ Differ 2016 Jul;96:127-131. [doi: 10.1016/j.paid.2016.02.085]
59. Randal C, Pratt D, Bucci S. Mindfulness and self-esteem: a systematic review. Mindfulness 2015 Apr 9;6(6):1366-1378. [doi: 10.1007/S12671-015-0407-6]
60. Farb N, Anderson A, Ravindran A, Hawley L, Irving J, Mancuso E, et al. Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy. J Consult Clin Psychol 2018 Feb;86(2):200-204. [doi: 10.1037/ccp0000266] [Medline: 29265831]
61. Katterman SN, Kleinman BM, Hood MM, Nackers LM, Corsica JA. Mindfulness meditation as an intervention for binge eating, emotional eating, and weight loss: a systematic review. Eat Behav 2014 Apr;15(2):197-204. [doi: 10.1016/j.eatbeh.2014.01.005] [Medline: 24854804]
62. Mindfulness based cognitive therapy and emotion focused therapy as treatments for major depression. Gena Fawns. 2013. URL: https://genafawns.com.au/wp-content/uploads/2016/01/Mindfulness_and_EFT_for_Depression_2013.pdf [accessed 2023-02-01]
63. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
64. Person SM. A two-study investigation into the link between rumination and night eating, and symptom improvement following a mindfulness-based intervention. Ontario Lakehead University. URL: https://knowledgecommons.lakeheadu.ca/handle/2453/4741 [accessed 2023-04-09]

65. Schanche E, Vøllestad J, Visted E, Svendsen JL, Osnes B, Binder PE, et al. The effects of mindfulness-based cognitive therapy on risk and protective factors of depressive relapse - a randomized wait-list controlled trial. BMC Psychol 2020 Jun 05;8(1):57 [FREE Full text] [doi: 10.1186/s40359-020-00417-1] [Medline: 32503649]
66. Rech DL, Schmidt KE, Rudnicki T, Schmidt MM. Techniques for managing the emotion of anger: a systematic review. Stud Res Psychol 2022 Apr 25;22(1):292-307 [FREE Full text] [doi: 10.12957/epp.2022.66485]
67. Moloud R, Saeed Y, Mahmonir H, Rasool GA. Cognitive-behavioral group therapy in major depressive disorder with focus on self-esteem and optimism: an interventional study. BMC Psychiatry 2022 Apr 28;22(1):299 [FREE Full text] [doi: 10.1186/s12888-022-03918-y] [Medline: 35484528]
68. Dobkin RD, Mann SL, Interian A, Gara MA, Menza M. Cognitive behavioral therapy improves diverse profiles of depressive symptoms in Parkinson's disease. Int J Geriatr Psychiatry 2019 May 04;34(5):722-729 [FREE Full text] [doi: 10.1002/gps.5077] [Medline: 30714202]
69. Santos CH. Efficacy of procedural cognitive therapy and behavioral activation in the treatment of Major Depressive Disorder: a randomized clinical trial. Federal University of Bahia. 2017. URL: https://sucupira-legado.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=6013054 [accessed 2023-04-10]
70. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018 Feb 30;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
71. Korrelboom K, Maarsingh M, Huijbrechts I. Competitive memory training (COMET) for treating low self-esteem in patients with depressive disorders: a randomized clinical trial. Depress Anxiety 2012 Feb;29(2):102-110. [doi: 10.1002/da.20921] [Medline: 22495940]
72. Wang X, Zhou X, Yang H. Guided self-help behavioral activation intervention for geriatric depression: protocol for pilot randomized controlled trial. JMIR Res Protoc 2020 Sep 24;9(9):e18259 [FREE Full text] [doi: 10.2196/18259] [Medline: 32518060]
73. Alves KI, Bonvicini CR. The role of behavioral activation in the management of depressive symptoms. Res Soc Develop 2022 Jan 04;11(1):e15311123711 [FREE Full text] [doi: 10.33448/rsd-v11i1.23711]
74. Wisenthal A, Krupa T, Kirsh B, Lysaght R. Insights into cognitive work hardening for return-to-work following depression: qualitative findings from an intervention study. Work 2019;62(4):599-613 [FREE Full text] [doi: 10.3233/WOR-192893] [Medline: 31104044]
75. Berman MI, Morton SN, Hegel MT. Uncontrolled pilot study of an acceptance and commitment therapy and health at every size intervention for obese, depressed women: accept yourself!. Psychotherapy (Chic) 2016 Dec;53(4):462-467 [FREE Full text] [doi: 10.1037/pst0000083] [Medline: 27797546]

Abbreviations

JBI: Joanna Briggs Institute
PCC: population, concept, and context
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by T de Azevedo Cardoso; submitted 04.07.24; peer-reviewed by MJ Norton, E Oliveira; comments to author 01.10.24; revised version received 21.10.24; accepted 19.12.24; published 06.03.25.

Please cite as:
de Campos Tibúrcio P, Maria Marcheti P, Miori Pascon D, Montebello Junior MA, Alzete de Lima M, Fernandes CS, Samarina Vilaça de Brito Santos C, Socorro de Sousa Nóbrega MDP
Therapeutic Guidelines for the Self-Management of Major Depressive Disorder: Scoping Review
Interact J Med Res 2025;14:e63959
URL: https://www.i-jmr.org/2025/1/e63959
doi:10.2196/63959
PMID:40053745

Review

Comparison of Outcomes Between Staged and Same-Day Circumferential Spinal Fusion for Adult Spinal Deformity: Systematic Review and Meta-Analysis

Mert Marcel Dagli[1], MD; Ryan William Turlip[1], BA; Felix C Oettl[2,3], MD; Mohamed Emara[4]; Jaskeerat Gujral[1]; Daksh Chauhan[1], BS; Hasan S Ahmad[1], MD; Gabrielle Santangelo[1], MD; Connor Wathen[1], MD; Yohannes Ghenbot[1], MD; John D Arena[1], MD; Joshua L Golubovsky[1], MD; Ben J Gu[1], MD; John H Shin[1], MD; Jang Won Yoon[1], MD, MSc; Ali K Ozturk[1], MD; William C Welch[1], MD

[1]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Hospital for Special Surgery, New York, NY, United States

[3]Department of Orthopedic Surgery, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

[4]College of Medicine, University of Sharjah, Sharjah, United Arab Emirates

Corresponding Author:
Mert Marcel Dagli, MD
Department of Neurosurgery
Perelman School of Medicine
University of Pennsylvania
801 Spruce Street
Philadelphia, PA, 19107
United States
Phone: 1 2672306493
Email: Marcel.Dagli@Pennmedicine.upenn.edu

Abstract

Background: Adult spinal deformity (ASD) is a prevalent condition often treated with circumferential spinal fusion (CF), which can be performed as staged or same-day procedures. However, evidence guiding the choice between these approaches is lacking.

Objective: This study aims to compare patient outcomes following staged and same-day CF for ASD.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, a comprehensive literature search was conducted in PubMed, MEDLINE, Embase, Cochrane CENTRAL, Web of Science, and Scopus. Eligibility criteria included studies comparing outcomes following staged and same-day CF in adults with ASD. Searches were exported to Covidence, and records were deduplicated automatically. Title and abstract screening, full-text review, and data extraction were performed by two independent reviewers, with all conflicts being resolved by a third reviewer. A meta-analysis was conducted for outcomes reported in 3 or more studies.

Results: Seven studies with 741 patients undergoing CF for ASD were included in the review (staged: n=331, 44.7% and same-day: n=410, 55.3%). Four studies that had comparable outcomes were merged for the quantitative meta-analysis and split based on observed measures. The meta-analysis revealed significantly shorter hospital length of stay (mean difference 3.98, 95% CI 2.23-5.72 days; $P<.001$) for same-day CF. Three studies compared the operative time between staged and same-day CF, with all reporting a lower mean operative time for same-day CF (mean between 291-479, SD 129 minutes) compared to staged CF (mean between 426-541, SD 124 minutes); however, inconsistent reporting of mean and SD made quantitative analyses unattainable. Of the 4 studies that compared estimated blood loss (EBL) in the relevant groups, 3 presented a lower EBL (mean between 412-1127, SD 954 mL) in same-day surgery compared to staged surgery (mean between 642, SD 550 to 1351, SD 869 mL). Both studies that reported intra- and postoperative adverse events showed more intraoperative adverse events in staged CF (10.9% and 13.6%, respectively) compared to same-day CF (9.1% and 3.6%, respectively). Four studies measuring any perioperative adverse events showed a higher incidence of adverse events in staged CF than all studies combined. However, quantitative analysis of EBL, intraoperative adverse events, and perioperative adverse events found no statistically significant difference. Postoperative adverse events, reoperation, infection rates, and readmission rates showed inconsistent findings between studies. Data quality assessment revealed a moderate degree of bias for all included studies.

Conclusions: Same-day CF may offer shorter operating time and hospital stay compared to staged CF for ASD. However, there was marked heterogeneity in perioperative outcomes reporting, and continuous variables were inconsistently presented. This underscored the need for standardized reporting of clinical variables and patient-reported outcomes and higher evidence of randomized controlled trials to elucidate the clinical superiority of either approach.

Trial Registration: PROSPERO CRD42022339764; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=339764

International Registered Report Identifier (IRRID): RR2-10.2196/42331

(Interact J Med Res 2025;14:e67290) doi:10.2196/67290

KEYWORDS

adults; circumferential fusion; scoliosis; spinal curvature; spinal fusion; spinal deformity; intraoperative; postoperative; perioperative; systematic reviews; meta-analysis; PRISMA

Introduction

Adult spinal deformity (ASD) is defined as abnormal curvature of the spine and is becoming increasingly prevalent, affecting up to 68% of the older adult population [1,2]. ASD is a complex spectrum of spinal pathology, including deformities such as lordosis, kyphosis, or scoliosis of the lumbar and thoracic spinal column. Although untreated adolescent ASD does occur, it typically presents in patients older than 60 years due to factors such as age-related spinal degeneration or reduced bone density [1,3].

Individuals with ASD can undergo expectant (observation alone), nonoperative, or operative therapies. At present, there is no high-quality evidence to support the decisions surgeons and patients face in treatment selection [4]. In past years, pain management and physical therapy were the preferred treatment options for ASD due to the high risk of adverse events, prolonged recovery time, and financial burden associated with surgical intervention [1,5]. If nonsurgical approaches fail to improve patients' quality of life, surgical intervention is often considered. Multicenter retrospective cohort studies previously showed an improvement in patient-reported outcomes following the surgical management of ASD [3,6]. Indications for surgery include (1) progressive curvature of the spine with sagittal or coronal imbalance, (2) significant loss of pulmonary function caused by the misalignment and deformity, and (3) loss of function due to pain associated with spinal curvature [7-10]. These are weighted against the patient comorbidities and risks of operation [11].

Long-segment surgical management by circumferential spinal fusion (CF) has increased in popularity due to its added stability granted by both anterior and posterior fixation of the spinal column [12]. CF attempts to remedy the limitations of lateral approaches alone, such as the need for intraoperative patient repositioning, which increases operative time and puts the patient at risk for adverse events due to longer time under anesthesia [13-15]. ASD can be treated by CF in 2 primary ways: staged and same day. Staged fusions occur on 2 distinct operative days, while same-day fusions are completed within a single session. Staging is largely determined by surgeon preference and case complexity, which can cause variability in the clinical management of ASD. The preference to treat with or without staging does not necessitate a gold-standard treatment for a given case complexity but rather can depend on surgical training differences and hospital administration pressures. To our knowledge, there has not been a review of published literature on staging in CF. This systematic review and meta-analysis aims to assess and quantify the patient outcomes after staged and same-day CF for ASD to guide operative decision-making and patient selection.

Methods

Guidelines, Protocol, and Registration

The design and reporting of this study were supported by the following guidelines: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1) and PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) [16,17]. In accordance with PRISMA-P guidelines, the protocol of the systematic review was registered on the PROSPERO (CRD42022339764) and disseminated through *JMIR Research Protocols* (PRR1-10.2196/4233), with the protocol being published before any data were collected [12]. There were no deviations from the protocol.

Eligibility Criteria

The Population, Intervention, Comparison, Outcome (PICO) framework was used to formulate inclusion criteria.

- Population: Adults with ASD
- Intervention: Staged CF surgery
- Comparison: Same-day CF surgery
- Outcomes: Perioperative outcomes (estimated blood loss [EBL], operative time, and length of hospital stay), adverse events, infection rates, and hospital readmissions or reoperations

Studies that do not differ in surgical timing (staged vs same day), nonhuman or adolescent patient populations, reviews, conference abstracts, single-case studies, or technical notes were excluded from the analysis. Further, only studies originally published in English were considered.

Search Strategy

Databases explored included PubMed, MEDLINE, Embase, Cochrane CENTRAL, Web of Science, and Scopus. A literature search was conducted in accordance with the PRISMA guidelines on August 2, 2023. We used a complex search string that was modified and fitted to the unique search functions of each queried database (Multimedia Appendix 2). Additional

searching through gray literature and reference lists was conducted to identify studies not initially captured by the database query.

Data Selection and Extraction

Studies and full text were screened, data were extracted using Covidence (Veritas Health Innovation), and duplicates were automatically removed by the software [18]. Titles and abstracts were first screened independently by 2 reviewers (FCO and ME). Next, the full text of each paper was assessed by 2 reviewers (FCO and ME) to determine the eligibility of the studies. At both stages, a third reviewer (MMD) resolved any conflicts. The following data were extracted by two authors (FCO and ME) with a third (MMD) resolving any conflicts: authors; publication year; location; number of patients; age; study type; population details; surgery details; and results, including intraoperative adverse events, postoperative adverse events, postoperative infection, perioperative adverse events, hospital length of stay (LOS), intensive care unit (ICU) LOS, reoperation, readmission, and patient-reported outcomes.

Data Quality

The ROBINS-I (Cochrane) tool was used to assess the risk of bias in the included nonrandomized studies, covering bias due to confounding variables, patient selection, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of reported results [19]. Two reviewers (FCO and JG) independently scored all domains, with a third reviewer (MMD) resolving any conflicts. Robvis was used for figure generation [20].

Data Synthesis

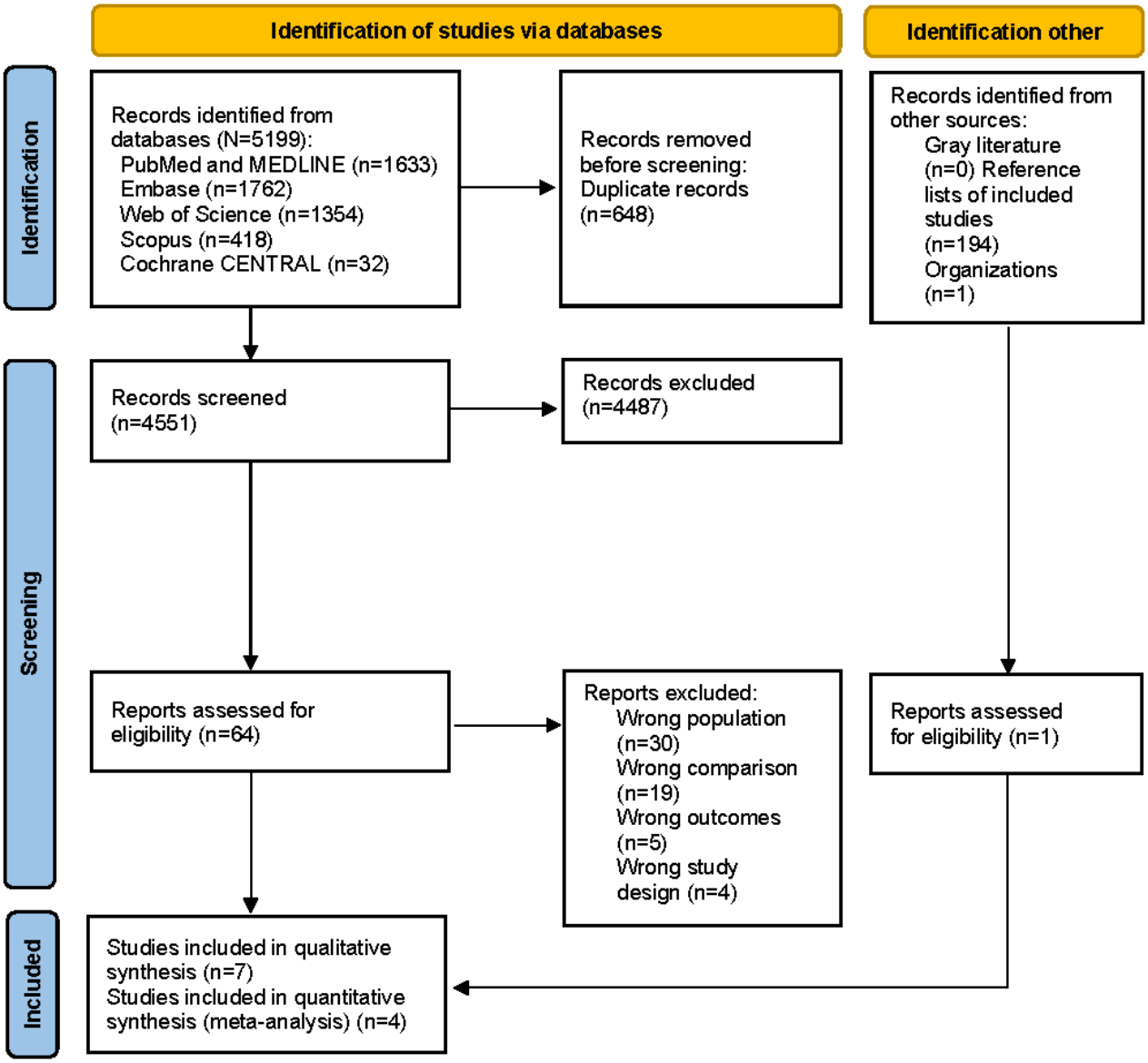
Studies with comparable outcomes were merged for the quantitative meta-analysis and split based on observed measures. After conducting a qualitative evaluation, we determined that there were sufficient data to perform a meta-analysis. RevMan (version 8.4; Cochrane) using random-effects modeling was used for all quantitative analyses. Mean differences for continuous variables (surgical time, EBL, hospital LOS, and ICU LOS) and odds ratios for categorical variables (intraoperative adverse events, postoperative adverse events, postoperative infection, any adverse events, readmission, and reoperation) were the end points of the meta-analysis.

Results

Study Identification

In our search (Figure 1), we identified 5199 unique studies by searching PubMed, MEDLINE, Embase, Cochrane CENTRAL, Web of Science, and Scopus, which were included for abstract screening, of which 64 were forwarded for full-text screening. After full-text review, 7 original studies were included in the data extraction process and 4 of them were in the quantitative analysis [21-27]. Studies were excluded during the full-text review for the following reasons: wrong comparator (n=19, 29.7%), wrong patient population (n=19, 29.7%), pediatric population (n=11, 17.2%), wrong outcomes reported (n=5, 7.8%), and wrong study design (n=4, 6.3%).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) flow diagram.



Baseline Parameters

Six of the studies included were conducted in the United States [21-25,27], while 1 study was conducted in Japan [26]. The included studies describe a total of 741 patients undergoing either staged (n=331, 44.7%) or same-day (n=410, 55.3%) CF for ASD: 297 (40.1%) patients underwent anterior lumbar interbody fusion, 408 (55.1%) patients underwent lateral interbody fusion (extreme lateral interbody fusion, direct lateral interbody fusion, and lateral lumbar interbody fusion), 54 (7%) patients underwent either lateral lumbar interbody fusion or transforaminal lumbar interbody fusion, and all 741 patients underwent posterior spinal fixation. The average number of fused vertebrae across studies ranged from 4.4 to 10 (SD 3.9). The largest variation between groups within a study ranged from 7.3 (SD 3.1) in the same-day group versus 10 (SD 3.9) in the staged group. The combined mean vertebrae fused for staged and same-day CF was 7.54 (SD 2.41) and 6.62 (SD 2.40), respectively. The follow-up period over all included studies ranged from 1 to more than 36 months, and the average age of patients ranged from 58.8 (SD 9.0) to 72.3 years. The patients included in the study of Masuda et al [26] and Albayar et al [21] were inverse probability weighted to facilitate the comparison of spinal deformity and control for any differences between the groups (Table 1).

Table 1. Patient demographics and significant results of included studies.

Authors (year)	Country	Patients	Age	Study type	Population details and differences	Surgery details	Results
Albayar et al (2023) [21]	United States	• n=44 • n=56	• Mean (SD): 58.8 (9.0) years • Mean (SD): 62.0 (11.9) years	• Retrospective cohort study, inverse probability weighted	• Patients aged >18 years at the time of surgery and diagnoses of ASDa undergoing (ALIFb), and open posterior lumbar or thoracolumbar (PSFc)	• Staged ALIF, and open posterior lumbar or thoracolumbar PSF, post. Vertebrae fused: mean 10 (SD 3.9) • Same-day ALIF, and open posterior lumbar or thoracolumbar PSF, post. Vertebrae fused: mean 7.3 (SD 3.1)	• Staged: EBLd mean 1351.7 (SD 869) mL, LOSe mean 10.5 (SD 5) days, IOAEf (n=6), POAEg (n=30), reoperation (n=10), POIh (n=5), and readmission (n=10). • Same day: EBL mean 1127.6 (SD 945.4), LOS mean 6.2 (SD 3.1) days, IOAE (n=2), POAE (n=30), reoperation (n=8), POI (n=1), and readmission (n=8).
Anand et al (2014) [22]	United States	• n=37 • n=13	• Mean (range): 61 (20-85) years	• Retrospective cohort study	• Patients with adult idiopathic scoliosis corrections undergoing cMISSi, Cobb angle of greater than 30 but less than 75 degrees	• Staged DLIFj and L5-S1 XLIFk followed by PSF. Vertebrae fused: mean 7 (range 4-15) • Same-day DLIF and L5-S1 XLIF followed by PSF. Vertebrae fused: mean 7 (range 4-15)	• Staged: EBL mean 763 (range 25-2500) mL and ORl time mean 482 (range 83-546) minutes. • Same day: EBL mean 613 (range 150-1500) ml and OR time mean 351 (range 176-510) minutes.
Anand et al (2013) [23]	United States	• n=36 • n=35	• Mean (range): 64 (20-84) years	• Retrospective cohort study	• Adults with scoliosis undergoing cMISS, two or more levels	• Stage combination of DLIF and XLIF with PSF. Vertebrae fused: mean 4.4 • Same-day combination of DLIF and XLIF with PSF. Vertebrae fused: mean 4.4	• Staged: EBL 671 mL and OR time 426 minutes. • Same day: EBL 412 mL and OR time 291 minutes.
Arzeno et al (2019) [24]	United States	• n=45 • n=47	• Mean (95% CI): 68 (61-78) years • Mean (95% CI): 68 (62-72) years	• Retrospective cohort study	• Patients with ASD, undergoing anterior (including lateral and anterolateral approaches) and PSF of at least 5 levels • Groups differ in approach, Ponte osteotomy, 3-column osteotomy, O-arm, neuromonitoring, decompression, number of posterior levels, fused; number. of osteotomy levels, mean; and number of decompression levels	• Staged circumferential spinal fusion (anterior, posterior), Ponte osteotomy (n=39), 3-column osteotomy (n=7), and decompression (n=34). Vertebrae fused: mean 8 (95% CI 5-9) • Same-day circumferential spinal fusion (anterior, posterior), Ponte osteotomy (n=24), 3-column osteotomy (n=1), and decompression (n=16). Vertebrae fused: mean 9 (95% CI 9-9)	• Staged: LOS mean 9 days, reoperation (n=5), readmission (n=1), POI (n=2), and PEAEm (n=12). • Same day: LOS mean 6 days, reoperation (n=7), readmission (n=6), POI (n=3), and PEAE (n=7).

Authors (year)	Country	Patients	Age	Study type	Population details and differences	Surgery details	Results
Harris et al (2021) [25]	United States	• n=41 • n=46	• Mean (SD): 62.0 (8.3) years • Mean (SD): 60.0 (13.0) years	• Retrospective cohort study	• Patients with ASD who underwent long PSF (more than 5 levels fused, with fusion to the pelvis) • Groups differ in previous spine surgery, scoliosis or kyphosis, pseudarthrosis, and pelvic incidence	• Staged circumferential -ALIF and PSF. Vertebrae fused: mean 8.7 (SD 0.48) • Same-day circumferential – ALIF and PSF. Vertebrae fused: mean 7.4 (SD 2.4)	• Staged: ODIn mean 45 (SD 17) and SRS-22ro mean 2.8 (SD 0.6) • Same day: ODI mean 48 (SD 15) and SRS-22r mean 2.8 (SD 0.6)
Masuda et al (2023) [26]	Japan	• n=101 • n=186	• Mean (SD): 72.2 (7.3) years • Mean (SD): 72.4 (8.0) years	• Retrospective cohort study, propensity score weighted	• Patients with ASD, ≥4 fused levels and at least 1 level using LLIFp, and presence of at least 1 spinal deformity marker: scoliosis Cobb angle≥20°, sagittal vertical axis≥5 cm, pelvic tilt≥25°, pelvic incidence minus lumbar lordosis angle≥10°, and thoracic kyphosis≥60°	• Staged circumferential-LLIF and PSF. Vertebrae fused: mean 7.7 (SD 2.3) • Same day circumferential-LLIF and PSF. Vertebrae fused: mean 6.2 (SD 2.4)	• Staged: EBL mean 642.5 (SD 550.5) mL, OR time mean 541.3 (SD 124.1) minutes, LOS mean 42 (SD 25) days, IOAE (n=11), POAE (n=11), reoperation (n=11), POI (n=4), and AAE[q] (n=22). • Same day: EBL 722.2 (SD 612.6) mL, OR time mean 479.9 (SD 128.5) minutes, LOS mean 34.1 (SD 18.2) days, IOAE (n=17), POAE (n=23), reoperation (n=19), POI (n=5), and AAE (n=40).
Than et al (2019) [27]	United States	• n=27 • n=27	• N/Ar • N/A	• Retrospective cohort study	• Patients with ASD, coronal Cobb angle >20, SVAs >5 cm, pelvic tilt >20, pelvic incidence–LLt >10, and thoracic kyphosis >60.	• Staged MISu LLIF or MIS TLIFv with PSF. Vertebrae fused: mean 5.4 • Same-day MIS LLIF or MIS TLIF with PSF. Vertebrae fused: mean 5.3	• Staged: Reoperation (n=4), POI (n=0), and PEAE (n=9). • Staged: Reoperation (n=7), readmission (n=1), POI (n=1), and PEAE (n=8).

[a]ASD: adult spinal deformity.

[b]ALIF: anterior lumbar interbody fusion.

[c]PSF: posterior spinal fusion.

[d]EBL: estimated blood loss.

[e]LOS: length of stay.

[f]IOAE: intraoperative adverse events

[g]POAE: postoperative adverse events.

[h]POI: postoperative infection.

[i]cMISS: circumferential minimally invasive spinal surgery.

[j]DLIF: direct lateral interbody fusion.

[k]XLIF: extreme lateral interbody fusion.

[l]OR: operating room.

[m]PEAE: perioperative adverse events.

[n]ODI: Oswestry Disability Index.

[o]SRS-22r: Scoliosis Research Society-22 revised.

[p]LLIF: lateral lumbar interbody fusion.

[q]AAE: any adverse events.

[r]N/A: not available.

[s]SVA: sagittal vertical axis.

[t]LL: lumbar lordosis.

[u]MIS: minimally invasive surgery.

[v]TLIF: transforaminal lumbar interbody fusion.

Quantitative Analysis

Of the included studies, 4 studies compared EBL in the relevant groups, with 3 studies presenting a lower EBL (mean between 412-1127, SD 954 mL) in same-day surgery compared to staged surgery (mean between 642, SD 550, and 1351, SD 869 mL) [21-24]. The meta-analysis shows a nonsignificant advantage for same-day surgery **(**Figure 2A [21,24,26]). Only 2 studies that measured EBL were included in the quantitative analysis because of inconsistencies in reporting, where some did not report variables as measures of variance, which made pooling in these instances not feasible. Three studies compared the operative time between staged and same-day CF, with all of them reporting a lower mean operative time for same-day CF (mean between 291-479 minutes) compared to staged CF (mean between 426-541 minutes) [22-24]. Just 1 group reported mean and SD for odds ratio time, thus restricting the potential for a quantitative analysis.

Three studies comparatively evaluated the hospital LOS [21,22,25]. All three studies consistently found that the mean LOS was less for same-day CF (mean between 6-34.1 days) in comparison to staged CF (mean between 9-42 days). The meta-analysis clearly presented a shorter LOS in patients undergoing same-day CF compared to staged CF (Figure 2B).

Two of the 7 studies compared intraoperative and postoperative adverse events between staged and same-day CF procedures [21,26]. Both Masuda et al [26] and Albayar et al [21] reported more intraoperative adverse events in staged CF (10.9% and 13.6%, respectively) compared to same-day CF (9.1% and 3.6%, respectively); however, the meta-analysis failed to show a statistically significant difference between the groups (Figure 3A [21-27]). Masuda et al [26] reported fewer postoperative adverse events in staged CF (10.9 vs 12.4%), while Albayar et al [21] presented a lower incidence in same-day CF (53.6 vs 68.2%), without a significant difference in the meta-analysis (Figure 3B). Four studies measured any perioperative adverse events [21,22,26,27]. The overarching analysis showed a higher incidence of adverse events in Staged CF over all studies; however, the meta-analysis did not show significance (Figure 3D).

Figure 2. Forest plots comparing (A) estimated blood loss and (B) hospital length of stay between patients who underwent staged or same-day circumferential spinal fusion for adult spinal deformity. IV: inverse variance.

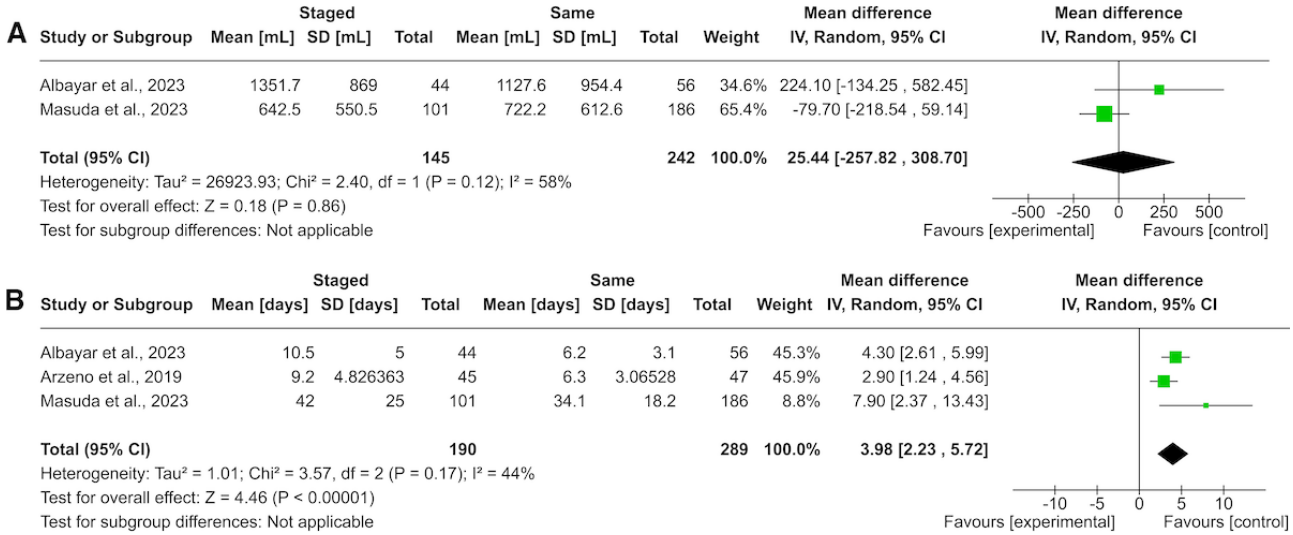
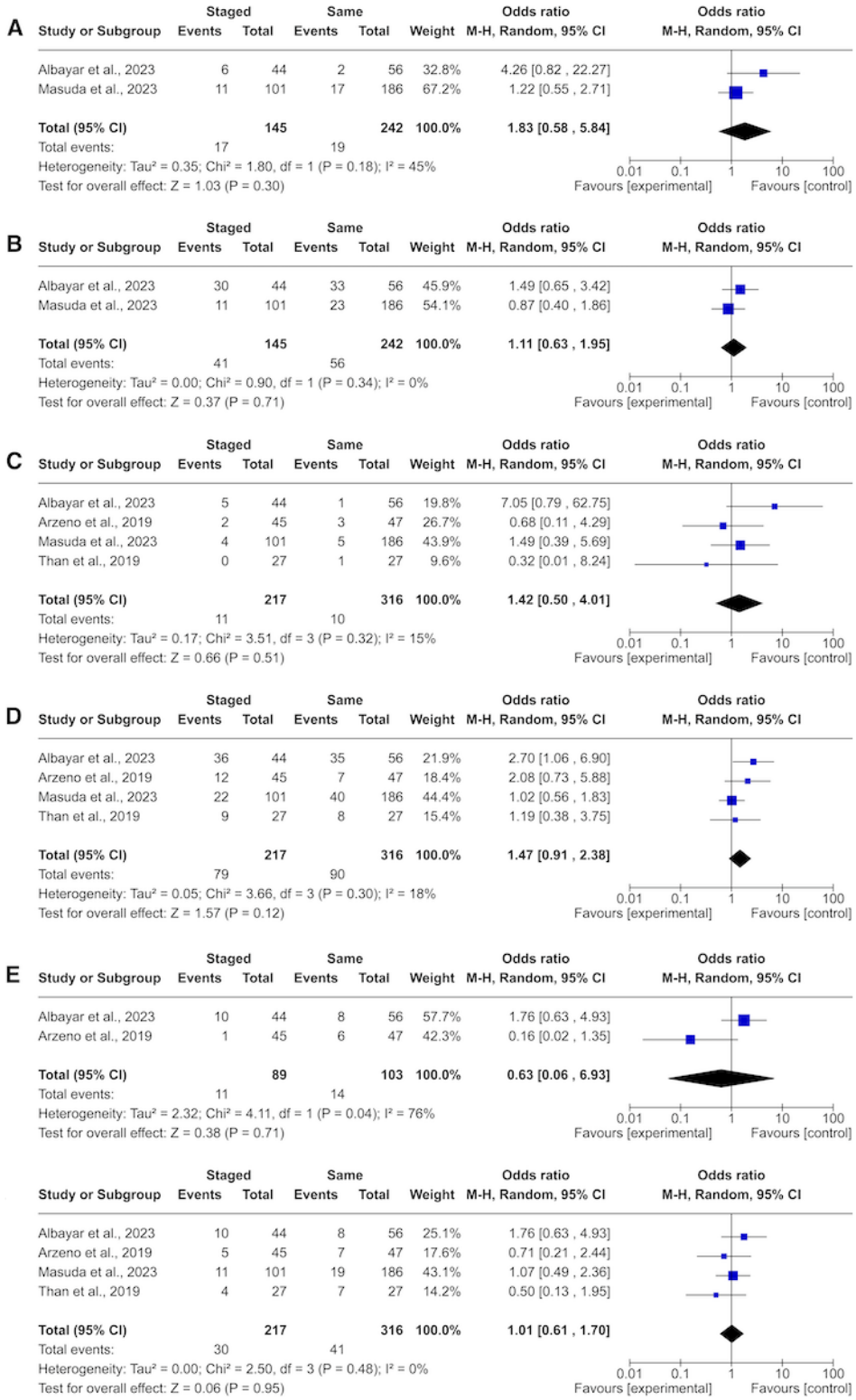


Figure 3. Forest plots comparing (A) intraoperative adverse event, (B) postoperative adverse event, (C) postoperative infection, (D) perioperative adverse event, (E) 30-day readmission, and (F) reoperation rates between patients who underwent staged or same-day circumferential spinal fusion for adult spinal deformity. M-H: Mantel-Haenszel.



Four of the included studies compared reoperation and postoperative infection rates between patient groups [21,22,26,27]. Albayar et al [21] and Masuda et al [26] showed a slightly lower postoperative infection in same-day CF versus staged CF (4.4 vs 6.4%, 2.6 vs 4%, and 1.7 vs 11.4% respectively), while Than et al [27] showed a lower postoperative infection in staged CF (0% vs 3.7%;Figure 3C). While Arzeno et al [24] and Than et al [27] reported a lower reoperation in patients undergoing staged CF (11.1 vs 14.9% and 14.8 vs 25.9%, respectively). Masuda et al [26] and Albayar et al [21] reported less reoperation in patients undergoing same-day CF versus staged CF (10.2 vs 10.9% and 14.2 vs 22.7% respectively). However, none of these differences in

reoperation reached statistical significance in either the original respective studies or our meta-analysis (Figure 3F).

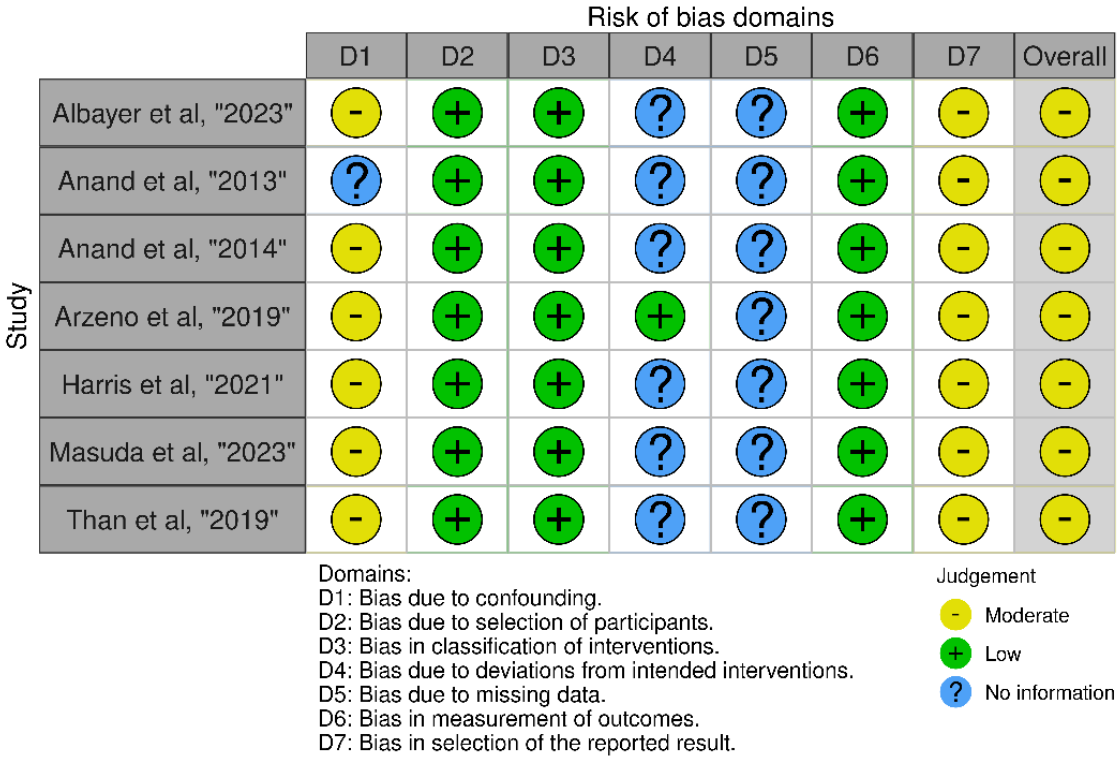
Readmission rates reported by Arzeno et al [24] and Albayar et al [21] demonstrated diverging results. Arzeno et al [24] found a lower readmission for staged CF versus same-day CF (2.2 vs 12.8%), while Albayar et al [21] reported a higher readmission in staged CF versus same-day CF (22.7 vs 14.3%). The meta-analysis did not indicate a conclusive result (Figure 3E).

Harris et al [25] are the only authors reporting on either the Oswestry Disability Index (ODI) or the Scoliosis Research Society Score, showing a better outcome measured by the ODI in same-day CF and no difference in Scoliosis Research Society Score.

Risk of Bias Analysis

There was a moderate degree of bias for all included studies (Figure 4 [21-27]).

Figure 4. ROBINS-I traffic light plot of included studies.



Discussion

Overview

In this systematic review and meta-analysis, we compared differences in perioperative outcome variables for individuals undergoing staged or same-day CF in 7 included studies. The meta-analysis revealed a statistically significant increase in hospital LOS in the staged group when compared to the same-day group. There were no statistically significant differences in EBL, intraoperative adverse event, postoperative adverse event, postoperative infection, perioperative adverse event, 30-day readmission, and reoperation rates between patients who underwent staged or same-day CF for ASD. There were marked differences between patient populations and the subsequent clinical outcomes that were reported in each observational study, which most did not adjust for. Our quantitative results are paradoxical as there were inconsistencies between studies in the staged and same-day subgroups for certain reported outcomes, while other findings had stronger conclusive evidence. Due to the apparent differences in consensus, the research included in the following discussion is organized according to the consistency of results.

Mixed Findings

Heterogeneity in meta-analyses grants the opportunity to examine variable factors that may be leading to the results. There were several reported outcomes that yielded inconclusive evidence in this study. Adverse event rates were reported by authors in several ways, with some splitting between intraoperative or postoperative adverse events, while others reported any adverse event throughout the perioperative course. Two studies reported more intraoperative adverse events in staged CF, but the meta-analysis failed to show a statistically significant difference. This can indicate that there is an increased operative risk when staging, but it can also be due to the generally sicker patients with more spinal deformity that is leading to increased intraoperative adverse events. Further, 1 study presented a lower postoperative adverse events rate during staged CF, while a different group found a higher rate in staged, indicating no difference in postoperative complications between the 2 approaches. When all perioperative adverse events were instead considered, every study reported a higher incidence in staged CF, and this result trended toward significance. Vastly different conclusions depending on the stage of surgery in which the adverse events are considered illuminate the need for greater consistency in reporting outcomes. However, the increased perioperative adverse events seen in staged CF, although not

statistically significant, could be clinically important in decision-making.

The Centers for Medicare & Medicaid Services targeted 30-day readmission rates as a source of unnecessary costs [28]. A study conducted by McCarthy et al [29] found the total hospital costs to surgically treat ASD averaged US $120,394, with primary surgery averaging US $103,143, and total readmission costs of US $67,262 for their cohort. The high costs associated with spine surgery make patient readmission and reoperation lucrative targets. There is profound heterogeneity in reoperation and readmission rates that depend on several factors, such as patient demographics, procedure types, and institutional factors. A recent systematic review found the 30-day readmission rate in spine surgery to be 4.2% and 7.4% [30]. In this review, 2 studies reported readmission with diverging results. Arzeno et al [24] found a lower readmission for staged CF, while Albayar et al [21] reported a higher readmission for same-day CF. Likewise, 4 studies reported reoperation with a split in consensus [21,24,26,27]. Given the small number of studies that measured reoperation and readmission, the results from our meta-analysis were inconclusive. The lack of statistically significant differences in the quantitative analysis combined with the heterogeneity in the qualitative review indicates that no differences in readmission and reoperation rates exist between the 2 groups.

It was also uncommon for the included studies to report specific reasons that reoperation or readmission occurred. As readmission depends on many variables including patient comorbidities, initial risk, or operative adverse events, it is crucial for authors to include these measures in future comparative studies for stronger subgroup analyses. Postoperative infection is also a driver leading to reoperation or readmission, as well as additional health care costs [31]. A recent meta-analysis found the pooled incidence of surgical site infection in 22,475 patients to be 3.1% [32]. Our included studies also reported a low incidence of postoperative infection; however, there were inconsistencies as 3 out of the 4 groups found a lower postoperative infection in same-day CF [21,24,26,27]. There are several risk factors for postoperative infection that may lead to heterogeneity. Farshad et al [33] found that intraoperative EBL was a risk factor for postoperative infection. Some studies found BMIs to be predictors of postoperative infection, although this result remains controversial [34-36]. The discrepancies seen in this review may stem from variations in surgical techniques, perioperative antibiotic prophylaxis, institutional differences, and patient-specific risk factors. The results of this study's meta-analysis do not support any difference between the 2 techniques in postoperation infection risk.

Associations of Operative Variables in CF Staging

For highly invasive spine surgeries, such as long-segmented CF for ASD, patients may require extended resource use [37]. Surgical operative time and LOS are critical metrics in evaluating the efficiency and resource use of staging in CF. Shorter surgical times are generally associated with reduced intraoperative adverse events, lower anesthesia-related risks, and decreased blood loss, thus contributing to improved patient safety and outcomes [15,38]. Peng et al [39] conducted a meta-analysis on correlations between operative time and postoperative infection, and they concluded that there was a 4-fold increase in postoperative infection risk for operations greater than 3 hours [39]. The studies included in our review all found a lower mean operative time for same-day CF.

A shorter LOS is also desirable as it minimizes health care resource use and lowers costs [40]. Further, early mobilization and discharge are associated with improved patient satisfaction and reduced psychological stress [41]. However, reducing LOS has also been associated with an increasing readmission rate at the population level [42]. Our meta-analysis found a significantly lower LOS for same-day CF compared to staged CF. This difference in LOS may be attributed to the cumulative effect of multiple hospital admissions, prolonged recovery periods between surgeries, and the need for additional preoperative preparation in staged CF. While the lower operative time and LOS results in patients undergoing same-day CF seem intuitive, it is important to interpret them cautiously, considering potential factors such as patient selection, discharge criteria, and institutional practices that may differ between studies. Moreover, while shorter operative time and LOS may reduce health care costs and improve resource allocation, they should not compromise patient safety or postoperative care quality.

The results of shorter LOS and operative time combined with the lack of conclusive differences between groups for adverse events, readmission, reoperation, and infection rates indicate that patient prognosis is similar when between staging preferences. The shorter time spent in the hospital and operating room is clinically important and contributes to improved patient safety and outcomes while reducing hospital resource use. However, the potential advantages of staging may outweigh these benefits and thus improve patient prognosis for more complex cases with a higher degree of relative deformity. Therefore, the results from this review and meta-analysis show preference toward same-day CF, while staging may largely depend on a patient's condition. Further studies investigating the impact of patient-specific factors, for example, varying degrees of deformity severity, on treatment selection and subsequent outcomes are required.

Limitations

There are several limitations to this study and the papers included in this review. First, the limited literature comparing staged and same-day CF resulted in only 7 studies being included, which reduced the statistical power of the meta-analysis as only 4 studies qualified for quantitative analysis. All the included studies are retrospective cohort studies with a level of evidence of III or II, and only 2 studies used inverse probability weighting to control for between-patient differences. Consequently, there was a moderate potential for bias, which limited our ability to reach generalizable conclusions. To elucidate the differences in surgical treatment options for CF, it will be necessary to investigate staging in level I or II randomized controlled trials (RCTs). Without RCTs, it is difficult to adjust for surgeon preference bias and the complexity of the case. Each study also demonstrated heterogeneity with respect to the patient outcomes reported,

further limiting the ability for robust statistical analysis. Continuous variables were generally poorly reported since many did not report measures of variance such as mean and SD, but rather mean or median and range (min-max or IQR). This made pooling in those instances not feasible, limiting the generalizability of results. Further, only 1 group presented patient-reported outcomes, such as ODI, which is a standardized clinical variable that can be useful for measuring subjective pain and disability. Novel studies should aim to investigate patient-reported outcomes and standardize the variables being reported, so future meta-analyses will have wider samples to yield more conclusive evidence for staging differences in CF. In doing so, the mixed results presented in this study can be further assessed to support wider implementation of either staged or same-day CF while integrating clinically significant metrics to evaluate patient outcomes.

Conclusions

Here, we present the first systematic review and quantitative meta-analysis on staging in CF to treat patients with ASD. Operative time and hospital LOS were significantly lower in same-day CF surgery, with EBL and perioperative adverse events also trending toward significance. However, there was heterogeneity in additional operative measures, such as intra- or postoperative adverse event rates, reoperation, and readmission, and no differences were found in our meta-analysis. Based on our results, it is suggested that same-day CF is advantageous as a potential time-save with patients spending less time in the operating room and hospital, potentially saving costs. However, it is still unclear whether either same-day or staged CF provides a clinical advantage for patient outcomes. Additional level I and II RCTs should be conducted to elucidate the associations between these variables and provide stronger evidence in favor of either approach.

Acknowledgments

This work was partially supported by the Catherine Sharpe Foundation.

Data Availability

All data generated or analyzed during this study are included in this published paper.

Authors' Contributions

WCW is the guarantor of the study. MMD and WCW led the conceptualization, data acquisition, analysis, and drafting and revision of the manuscript. RWT, JG, GS, DC, HSA, CAW, TG, JDA, JLG, and BJG contributed to data acquisition, analysis, and drafting. Independent review was performed by FCO, ME, and MMD. All authors contributed to the analysis, interpretation, and drafting. JHS, JWY, AKO, and WCW contributed critical guidance at all stages of the study. The manuscript was reviewed and edited, and its final version was approved by all authors.

Conflicts of Interest

In the past 36 months, AKO has received consulting fees from Medacta and Johnson and Johnson and has served as an E2M ad hoc reviewer for the Journal of Neurosurgery Publishing Group (JNS PG), with no relation to this work. Additionally, within the same period, JWY has received a grant from Johnson and Johnson; consulting fees from Medyssey, TrackX, Richard Wolf, and Johnson and Johnson; holds patents planned, issued, or pending with Kinesiometrics and MedCyclops; and has served in a leadership role on the Scientific Program Committee of the American Association of Neurological Surgeons (AANS)/Congress of Neurological Surgeons (CNS) Joint Section on Disorders of the Spine and Peripheral Nerves, with no relation to this work. All other authors report no conflict of interest.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 86 KB - ijmr_v14i1e67290_app1.pdf]

Multimedia Appendix 2
Supplemental search strings for query.
[DOCX File , 15 KB - ijmr_v14i1e67290_app2.docx]

References

1. Kim HJ, Yang JH, Chang D, Lenke LG, Suh SW, Nam Y, et al. Adult spinal deformity: a comprehensive review of current advances and future directions. Asian Spine J 2022;16(5):776-788 [FREE Full text] [doi: 10.31616/asj.2022.0376] [Medline: 36274246]
2. Cerpa M, Lenke LG, Fehlings MG, Shaffrey CI, Cheung KMC, Carreon LY. Evolution and advancement of adult spinal deformity research and clinical care: an overview of the Scoli-RISK-1 study. Global Spine J 2019;9:8S-14S [FREE Full text] [doi: 10.1177/2192568219828729] [Medline: 31157151]

3. Diebo BG, Shah NV, Boachie-Adjei O, Zhu F, Rothenfluh DA, Paulino CB, et al. Adult spinal deformity. Lancet 2019;394(10193):160-172. [doi: 10.1016/S0140-6736(19)31125-0] [Medline: 31305254]
4. Youssef JA, Orndorff DO, Patty CA, Scott MA, Price HL, Hamlin LF, et al. Current status of adult spinal deformity. Global Spine J 2013;3(1):51-62 [FREE Full text] [doi: 10.1055/s-0032-1326950] [Medline: 24436852]
5. Aebi M. The adult scoliosis. Eur Spine J 2005;14(10):925-948. [doi: 10.1007/s00586-005-1053-9] [Medline: 16328223]
6. Kim HJ, Chang D. Clinical and radiographic parameters for patients with adult spinal deformity. J Korean Med Assoc 2021;64(11):743-747. [doi: 10.5124/jkma.2021.64.11.743]
7. Bradford DS, Tay BK, Hu SS. Adult scoliosis: surgical indications, operative management, complications, and outcomes. Spine 1999;24(24):2617-2629. [doi: 10.1097/00007632-199912150-00009] [Medline: 10635525]
8. Glassman SD, Berven S, Kostuik J, Dimar JR, Horton WC, Bridwell K. Nonsurgical resource utilization in adult spinal deformity. Spine 2006;31(8):941-947. [doi: 10.1097/01.brs.0000209318.32148.8b] [Medline: 16622386]
9. Acaroglu E, Yavuz AC, Guler UO, Yuksel S, Yavuz Y, Domingo-Sabat M, et al. A decision analysis to identify the ideal treatment for adult spinal deformity: Is surgery better than non-surgical treatment in improving health-related quality of life and decreasing the disease burden? Eur Spine J 2016;25(8):2390-2400. [doi: 10.1007/s00586-016-4413-8] [Medline: 26821143]
10. Ames CP, Scheer JK, Lafage V, Smith JS, Bess S, Berven SH, et al. Adult spinal deformity: epidemiology, health impact, evaluation, and management. Spine Deform 2016;4(4):310-322. [doi: 10.1016/j.jspd.2015.12.009] [Medline: 27927522]
11. Good CR, Auerbach JD, O'Leary PT, Schuler TC. Adult spine deformity. Curr Rev Musculoskelet Med 2011;4(4):159-167 [FREE Full text] [doi: 10.1007/s12178-011-9101-z] [Medline: 22021017]
12. Dagli MM, Narang S, Malhotra K, Santangelo G, Wathen C, Ghenbot Y, et al. The differences between same-day and staged (circumferential) fusion surgery in adult spinal deformity: protocol for a systematic review. JMIR Res Protoc 2022;11(11):e42331 [FREE Full text] [doi: 10.2196/42331] [Medline: 36441570]
13. Thomas JA, Menezes C, Buckland AJ, Khajavi K, Ashayeri K, Braly BA, et al. Single-position circumferential lumbar spinal fusion: an overview of terminology, concepts, rationale and the current evidence base. Eur Spine J 2022;31(9):2167-2174. [doi: 10.1007/s00586-022-07229-4] [Medline: 35913621]
14. Kim BD, Hsu WK, de Oliveira GS, Saha S, Kim JYS. Operative duration as an independent risk factor for postoperative complications in single-level lumbar fusion: an analysis of 4588 surgical cases. Spine 2014;39(6):510-520. [doi: 10.1097/BRS.0000000000000163] [Medline: 24365901]
15. Hersey AE, Durand WM, Eltorai AEM, DePasse JM, Daniels AH. Longer operative time in elderly patients undergoing posterior lumbar fusion is independently associated with increased complication rate. Global Spine J 2019;9(2):179-184 [FREE Full text] [doi: 10.1177/2192568218789117] [Medline: 30984498]
16. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
18. Covidence Systematic Review Software, Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2025-01-15]
19. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
20. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
21. Albayar A, Santangelo G, Spadola M, Macaluso D, Ali ZS, Saifi C, et al. Comparison of staged vs same-day circumferential spinal fusions for adult spinal deformity. Int J Spine Surg 2023;17(6):843-855 [FREE Full text] [doi: 10.14444/8548] [Medline: 37827708]
22. Anand N, Baron EM, Khandehroo B. Is circumferential minimally invasive surgery effective in the treatment of moderate adult idiopathic scoliosis? Clin Orthop Relat Res 2014;472(6):1762-1768 [FREE Full text] [doi: 10.1007/s11999-014-3565-2] [Medline: 24658900]
23. Anand N, Baron EM, Khandehroo B, Kahwaty S. Long-term 2- to 5-year clinical and functional outcomes of minimally invasive surgery for adult scoliosis. Spine 2013;38(18):1566-1575. [doi: 10.1097/BRS.0b013e31829cb67a] [Medline: 23715025]
24. Arzeno AH, Koltsov J, Alamin TF, Cheng I, Wood KB, Hu SS. Short-term outcomes of staged versus same-day surgery for adult spinal deformity correction. Spine Deform 2019;7(5):796-803.e1. [doi: 10.1016/j.jspd.2018.12.008] [Medline: 31495481]
25. Harris AB, Puvanesarajah V, Raad M, Marrache M, Ren M, Skolasky RL, et al. How is staging of ALIF following posterior spinal arthrodesis to the pelvis related to functional improvement in patients with adult spinal deformity? Spine Deform 2021;9(4):1085-1091. [doi: 10.1007/s43390-020-00272-5] [Medline: 33464551]

26. Masuda S, Fujibayashi S, Takemoto M, Ota M, Onishi E, Odate S, et al. Association of two-staged surgery with systemic perioperative complications in lateral lumbar interbody fusion for adult spinal deformity: a propensity score-weighted study. Eur Spine J 2023;32(3):950-956. [doi: 10.1007/s00586-023-07539-1] [Medline: 36680618]
27. Than KD, Park P, Tran S, Mundis GM, Fu K, Uribe JS, et al. Analysis of complications with staged surgery for less invasive treatment of adult spinal deformity. World Neurosurg 2019;126:e1337-e1342. [doi: 10.1016/j.wneu.2019.03.090] [Medline: 30898739]
28. McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. Circulation 2015;131(20):1796-1803 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.010270] [Medline: 25986448]
29. McCarthy IM, Hostin RA, Ames CP, Kim HJ, Smith JS, Boachie-Adjei O, et al. Total hospital costs of surgical treatment for adult spinal deformity: an extended follow-up study. Spine J 2014;14(10):2326-2333. [doi: 10.1016/j.spinee.2014.01.032] [Medline: 24469004]
30. Bernatz JT, Anderson PA. Thirty-day readmission rates in spine surgery: systematic review and meta-analysis. Neurosurg Focus 2015;39(4):E7. [doi: 10.3171/2015.7.FOCUS1534] [Medline: 26424347]
31. Anderson PA, Savage JW, Vaccaro AR, Radcliff K, Arnold PM, Lawrence BD, et al. Prevention of surgical site infection in spine surgery. Neurosurgery 2017;80(3S):S114-S123. [doi: 10.1093/neuros/nyw066] [Medline: 28350942]
32. Zhou J, Wang R, Huo X, Xiong W, Kang L, Xue Y. Incidence of surgical site infection after spine surgery: a systematic review and meta-analysis. Spine 2020;45(3):208-216. [doi: 10.1097/BRS.0000000000003218] [Medline: 31464972]
33. Farshad M, Bauer DE, Wechsler C, Gerber C, Aichmair A. Risk factors for perioperative morbidity in spine surgeries of different complexities: a multivariate analysis of 1,009 consecutive patients. Spine J 2018;18(9):1625-1631 [FREE Full text] [doi: 10.1016/j.spinee.2018.02.003] [Medline: 29452285]
34. Abdallah DY, Jadaan MM, McCabe JP. Body mass index and risk of surgical site infection following spine surgery: a meta-analysis. Eur Spine J 2013;22(12):2800-2809 [FREE Full text] [doi: 10.1007/s00586-013-2890-6] [Medline: 23828507]
35. Blam OG, Vaccaro AR, Vanichkachorn JS, Albert TJ, Hilibrand AS, Minnich JM, et al. Risk factors for surgical site infection in the patient with spinal injury. Spine 2003;28(13):1475-1480. [doi: 10.1097/01.BRS.0000067109.23914.0A] [Medline: 12838110]
36. Capen DA, Calderone RR, Green A. Perioperative risk factors for wound infections after lower back fusions. Orthop Clin North Am 1996;27(1):83-86. [doi: 10.1016/s0030-5898(20)32052-6]
37. Reis RC, de Oliveira MF, Rotta JM, Botelho RV. Risk of complications in spine surgery: a prospective study. Open Orthop J 2015;9:20-25 [FREE Full text] [doi: 10.2174/1874325001509010020] [Medline: 25674185]
38. Findlay MC, Kim RB, Warner WS, Sherrod BA, Park S, Mazur MD, et al. Identification of an operative time threshold for substantially increased postoperative complications among elderly spine surgery patients. Global Spine J 2024;14(5):1532-1541 [FREE Full text] [doi: 10.1177/21925682221149390] [Medline: 36623932]
39. Peng XQ, Sun CG, Fei ZG, Zhou QJ. Risk factors for surgical site infection after spinal surgery: a systematic review and meta-analysis based on twenty-seven studies. World Neurosurg 2019;123:e318-e329. [doi: 10.1016/j.wneu.2018.11.158] [Medline: 30496932]
40. Stephens BF, Khan I, Chotai S, Sivaganesan A, Devin CJ. Drivers of cost in adult thoracolumbar spine deformity surgery. World Neurosurg 2018;118:e206-e211. [doi: 10.1016/j.wneu.2018.06.155] [Medline: 29966783]
41. Smith G, Chirieleison S, Levin J, Atli K, Winkelman R, Tanenbaum JE, et al. Impact of length of stay on HCAHPS scores following lumbar spine surgery. J Neurosurg Spine 2019;31(3):366-371. [doi: 10.3171/2019.3.SPINE181180] [Medline: 31151093]
42. Ansari SF, Yan H, Zou J, Worth RM, Barbaro NM. Hospital length of stay and readmission rate for neurosurgical patients. Neurosurgery 2018;82(2):173-181. [doi: 10.1093/neuros/nyx160] [Medline: 28402465]

Abbreviations

ASD: adult spinal deformity
CF: circumferential spinal fusion
EBL: estimated blood loss
ICU: intensive care unit
LOS: length of stay
ODI: Oswestry Disability Index
PICO: Population, Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
RCT: randomized controlled trial

Edited by T de Azevedo Cardoso; submitted 07.10.24; peer-reviewed by J Shen; comments to author 05.11.24; revised version received 31.12.24; accepted 07.01.25; published 06.03.25.

Please cite as:
Dagli MM, Turlip RW, Oettl FC, Emara M, Gujral J, Chauhan D, Ahmad HS, Santangelo G, Wathen C, Ghenbot Y, Arena JD, Golubovsky JL, Gu BJ, Shin JH, Yoon JW, Ozturk AK, Welch WC
Comparison of Outcomes Between Staged and Same-Day Circumferential Spinal Fusion for Adult Spinal Deformity: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e67290
URL: https://www.i-jmr.org/2025/1/e67290
doi:10.2196/67290
PMID:40053742

Review

Informatics Interventions for Maternal Morbidity: Scoping Review

Jill Inderstrodt[1,2,3], MPH, PhD; Julia C Stumpff[4], MSLIS; Rebecca C Smollen[5], MPH; Shreya Sridhar[6], MPH; Sarah A El-Azab[7], MS; Opeyemi Ojo[8], MPH; Brendan Bowns[9], MPH; David A Haggstrom[3,10,11], MAS, MD

[1]Department of Health Policy and Management, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States
[2]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States
[3]Center for Health Information and Communication, Richard L. Roudebush VA Medical Center, Indianapolis, IN, United States
[4]Ruth Lilly Medical Library, Indiana University School of Medicine, Indianapolis, IN, United States
[5]Department of Public Health, College of Health and Human Sciences, Purdue University, West Lafayette, IN, United States
[6]School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, United States
[7]Department of Health Policy and Management, University of Michigan, Ann Arbor, MI, United States
[8]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States
[9]Department of Prevention, Corktown Health, Detroit, MI, United States
[10]Center for Health Services Research, Regenstrief Institute, Indianapolis, IN, United States
[11]Division of General Internal Medicine and Geriatrics, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Jill Inderstrodt, MPH, PhD
Department of Health Policy and Management
Richard M. Fairbanks School of Public Health
Indiana University
1050 Wishard Blvd.
Indianapolis, IN, 46202
United States
Phone: 1 3176903316
Email: ji3@iu.edu

Abstract

Background: Women have been entering pregnancy less healthy than previous generations, placing them at increased risk for pregnancy complications. One approach to ensuring effective monitoring and treatment of at-risk women is designing technology-based interventions that prevent maternal morbidities and treat perinatal conditions.

Objective: This scoping review evaluates what informatics interventions have been designed and tested to prevent and treat maternal morbidity.

Methods: MEDLINE, Embase, and Cochrane Library were searched to identify relevant studies. The inclusion criteria were studies that tested a medical or clinical informatics intervention; enrolled adult women; and addressed preeclampsia, gestational diabetes mellitus (GDM), preterm birth, Centers for Disease Control and Prevention–defined severe maternal morbidity, or perinatal mental health conditions. Demographic, population, and intervention data were extracted to characterize the technologies, conditions, and populations addressed.

Results: A total of 80 studies were identified that met the inclusion criteria. Many of the studies tested for multiple conditions. Of these, 73% (60/82) of the technologies were tested for either GDM or perinatal mental health conditions, and 15% (12/82) were tested for preeclampsia. For technologies, 32% (28/87) of the technologies tested were smartphone or tablet applications, 26% (23/87) were telehealth interventions, and 14% (12/87) were remote monitoring technologies. Of the many outcomes measured by the studies, almost half (69/140, 49%) were patient physical or mental health outcomes.

Conclusions: Per this scoping review, most informatics interventions address three conditions: GDM, preeclampsia, and mental health. There may be opportunities to treat other potentially lethal conditions like postpartum hemorrhage using proven technologies such as mobile apps. Ample gaps in the literature exist concerning the use of informatics technologies aimed at maternal morbidity. There may be opportunities to use informatics for lesser-targeted conditions and populations.

(Interact J Med Res 2025;14:e64826) doi:10.2196/64826

KEYWORDS

scoping review; maternal morbidity; medical informatics; clinical informatics; mother; pregnant; perinatal; GDM; preeclampsia; maternity; gestational diabetes mellitus

Introduction

Women have been entering pregnancy less healthy than previous generations [1]. Maternal morbidities such as preeclampsia and gestational diabetes, and severe maternal morbidities such as postpartum hemorrhage all have implications for the long-term health of mothers well beyond the postpartum period. Preeclampsia is associated with cardiovascular disease for decades following pregnancy, including chronic hypertension, stroke, and age-adjusted overall mortality [2]. Gestational diabetes mellitus (GDM), a condition that in many cases resolves itself after delivery, can still affect mothers beyond the puerperium [3]. Severe maternal morbidity (SMM) encompasses myriad conditions, including acute myocardial infarction, eclampsia, and hemorrhage. Those who have experienced SMM are more likely to die at any point after delivery and into the decades beyond the postpartum period [4]. Preterm birth (PTB; delivering before 37 weeks gestation) is associated with long-term cardiovascular complications in the mother, including ischemic heart disease, stroke, and atherosclerosis [5].

Informatics interventions offer tools that can help to prevent perinatal health conditions that have long-term health consequences for mothers, monitor these conditions in the perinatal period so that mother and baby remain healthy, and follow mothers post partum to ensure that they continue to receive the health monitoring they need. Technology-based tools that can be used for these purposes include mobile apps, wearable technology, physician decision support, and telehealth, among others. As artificial intelligence (AI) technology becomes more common in health care, a range of interventions can be used to predict, diagnose, and treat maternal morbidities that lead to maternal complications, with the hope of reducing maternal morbidity and mortality [6]. While some patients are hesitant for their physicians to use AI to diagnose or treat medical problems, physicians have demonstrated growing acceptance and belief that AI will improve interactions with their patients [7]. As the landscape of health informatics shifts substantially with the advent of new technologies, it is important to assess what technologies have already been tested to prevent, diagnose, and treat common maternal morbidities so that gaps can be identified and addressed in future research [8].

In this scoping review, we seek to identify what informatics interventions have been designed and tested that address maternal morbidity. In addition, we aim to assess the extent, range, and nature of informatics interventions that have been tested to prevent, diagnose, and treat maternal morbidities known to have long-term health consequences for mothers. The review summarizes discrete populations addressed in the literature, types of informatics interventions tested in the literature, conditions addressed by these interventions, and outcomes measured.

Methods

Search Strategy

Our scoping review methodology was guided by Arksey and O'Malley's [9] framework, stages 1-5, and Joanna Briggs's *Manual for Scoping Reviews*. We asked what informatics interventions have been designed and tested for SMM. Using the PCC (Population, Concept, and Context) framework outlined in the *Manual for Scoping Reviews*, the population was adult women with SMM as defined by the Centers for Disease Control and Prevention (CDC), the concept was informatics interventions, and the context was prospective studies anywhere in the world. We developed a review protocol to guide the process. The literature searches were led by an expert information specialist (JCS) in consultation with the research team. Four electronic databases were searched from inception until June 6, 2022: MEDLINE and Embase (searched simultaneously on Ovid), Cochrane Library (Wiley), and Engineering Village (Elsevier). One database was searched from inception until June 7, 2022: IEEE Xplore (IEEE). The MEDLINE (Ovid) search was peer-reviewed by another expert librarian using the PRESS (Peer Review of Electronic Search Strategies) checklist and modified as required [10]. The search strategy was limited to the English language and encompassed all years of publication. Embase and MEDLINE citations were deduplicated in Ovid before exporting to Covidence online review software. The full search strategies for each database are publicly available in searchRxiv [11]. Lastly, review results are reported using PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [12]. This checklist can be found in Multimedia Appendix 1. Our protocol was registered retrospectively in Open Science Framework Registries [13].

Study Selection

The inclusion and exclusion criteria are described in Textbox 1. Studies were included that prospectively tested a medical informatics intervention on either a group of women or physicians who treat women. Study methodologies could include both quantitative and qualitative approaches. Only studies analyzing primary data to describe a health-related outcome were included. Studies were included that tested medical or clinical informatics interventions on human subjects to prevent, monitor, or treat maternal conditions that have long-term health consequences for mothers within and beyond the perinatal period. Conditions included preeclampsia or hypertensive disorders of pregnancy, GDM, PTB, new mental health diagnoses, and SMM as defined by the US CDC [14].

Textbox 1. Criteria for inclusion and exclusion.

Inclusion criteria

- Test a medical informatics intervention prospectively
- Medical or clinical informatics interventions
- Adult women or physicians who treat women
- Preeclampsia, gestational diabetes mellitus, preterm birth, perinatal mental health diagnoses, severe maternal morbidity as defined by the Centers for Disease Control and Prevention (21 indicators)
- Physical health, patient-centered outcomes, mental health, health behavior, health knowledge or attitudes, health care use, quality of care outcomes

Exclusion criteria

- Systematic reviews, scoping reviews, literature reviews, opinion pieces, commentaries, proposals, reports, conference papers
- Pilot/feasibility studies
- Bioinformatics studies, including scans, ultrasounds, biomarkers, and predictive algorithms
- Adolescents
- Preexisting conditions: type 1 or 2 diabetes, chronic hypertension, preexisting mental health conditions
- Feasibility, acceptability, user experience

Systematic reviews, scoping reviews, and literature reviews were excluded together with opinion pieces, commentaries, proposals, reports, or gray literature. Studies were also excluded if they focused primarily on bioinformatics, effectively eliminating imaging (eg, ultrasounds), biomarkers, and predictive algorithm development.

Article references were entered into review software for screening and data extraction. The first and senior authors conducted an initial title, abstract, and keyword screening. If any author recommended inclusion, then the article underwent full-text screening. All authors except for the information specialist (JCS) then conducted a full-text screening, with two team members reviewing all articles. Studies with two votes to include were moved forward for extraction; those with two votes to exclude were excluded from the study. During the full-text screening, Cohen κ was used in the initial training to gauge agreement until reviewers reached a κ of 0.80. After training, disagreements were resolved by the first author.

Data Extraction

A subset of four authors extracted data using Covidence review software. Reviewers were given an initial set of 10 articles and met with the first author for consensus before reviewing the remaining articles. All studies were reviewed for extraction by the first author and one additional author. Most data extraction categories were developed a priori, and selection options were developed by the first and senior authors after reviewing all articles for trends. The following data categories were added after the protocol had been developed: aim of the study, years of data collection, population inclusion and exclusion criteria, and total number of participants. These categories were added to the data extraction form iteratively to provide more context to the reviewed studies.

Demographic data collected for each article by reviewers included title of article, year published, lead author surname, and country in which the study was conducted. The following characteristics of included studies were extracted: informatics technology tested, health problem, aim of study, study design, years of data collection, population description, study participation inclusion criteria, study participation exclusion criteria, total number of participants, and health outcome. Reviewers were given the option of selecting multiple answers for each study characteristic. An open-ended "other" option was available for the following study characteristics: informatics technology, health problem, study design, and outcome. Following extraction, the first and senior authors also extracted data on whether a significantly positive result was found in the studies, as well as racial diversity in participant samples for those studies conducted in the United States.

Data were synthesized by tallying the options from each category (study design, region, technology, health outcome, study population, study outcome, and results) and calculating the corresponding frequencies for each option.

Results

Overview

The PRISMA flow diagram for this scoping review can be found in Figure 1. A total of 3799 records were identified for title and abstract screening; 652 records underwent full-text screening for eligibility. Of these, 572 were excluded. Common reasons for exclusion were study design (eg, retrospective, systematic or other review, narrative), type of study outcome (user experience), or the wrong intervention (eg, machine learning algorithms or predictive models not tested with patients or providers). A total of 80 studies underwent full data extraction. A list of the studies and their characteristics can be found in Table 1.

Frequencies for synthesized data can be found in Table 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.

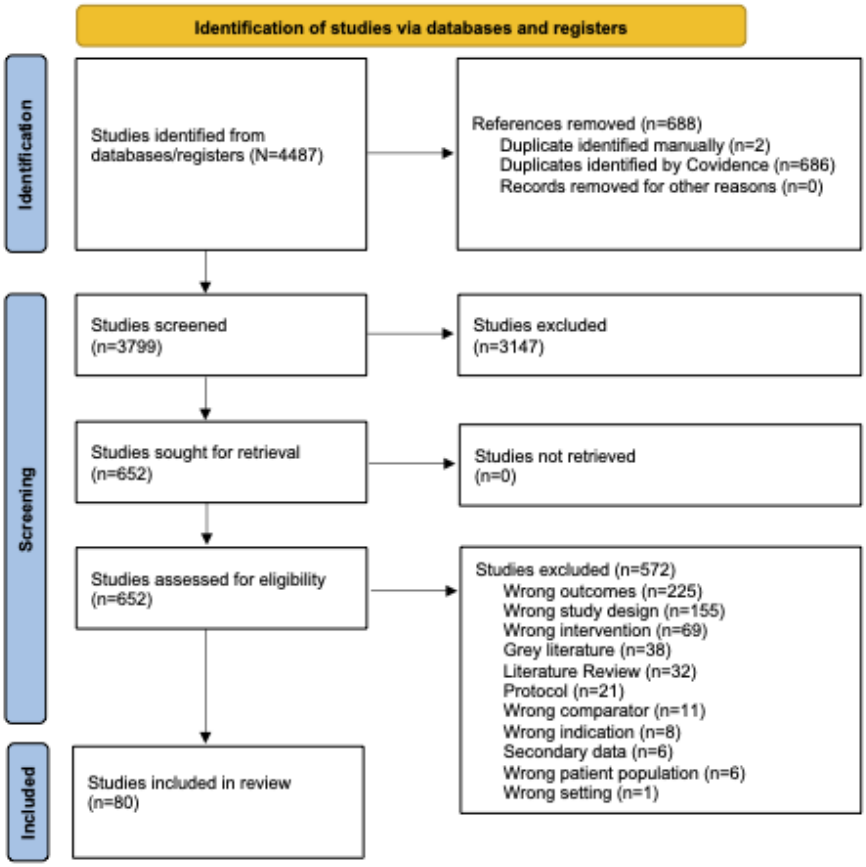


Table 1. Studies selected for extraction.

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Abbaspoor et al [15] 2020	Iran	• SMS text message	GDM[a]	RCT[b]	Women	100	• Medical • Health behavior
Abejirinde et al [16] 2019	Ghana	• Smartphone/tablet application • DSSc	HDP[d], GDM	Nonrandomized experimental	Women	940	• Quality of care
Al-Ofi et al [17] 2019	Saudi Arabia	• Remote monitoring • Telehealth	GDM	RCT	Women	57	• Medical • Health behavior
Arias et al [18] 2022	United States	• Telehealth	HDP	Cohort	Women	1579	• Health behavior • Health care use • Quality of care
Bartholomew et al [19] 2015	United States	• Remote monitoring • Smartphone/tablet application • Web program	GDM	RCT	Women	74	• Number of readings
Baumel et al [20] 2018	United States	• Smartphone/tablet application • Web program	Mental health condition	Nonrandomized experimental	Women	19	• Mental health
Bellad et al [21] 2020	India	• Web application	HDP	RCT	Providers	14,783	• Medical
Borgen et al [22] 2019	Norway	• Smartphone/tablet application	GDM	RCT	Women	158	• Medical • Health behavior
Carlisle et al [23] 2022	United Kingdom	• DSS	Mental health condition, PTB[e]	RCT	Women	300	• Mental health
Carroll et al [24] 2013	United States	• DSS	Mental health condition	RCT	Providers	48	• Quality of care
Chan et al [25] 2019	China	• Smartphone/tablet application	Mental health condition	RCT	Women	660	• Mental health
Chappell et al [26] 2022	United Kingdom	• Remote monitoring	HDP	RCT	Women	454	• Medical
Cheung et al [27] 2019	Australia	• SMS text messaging • Wearable	GDM	RCT	Women	60	• Medical • Number of readings • Health behavior
Dennis et al [28] 2009	Canada	• Telehealth	Mental health condition	RCT	Women	846	• Medical • Mental health
Dennis et al [29] 2020	United States	• Telehealth	Mental health condition	RCT	Women	249	• Medical • Mental health • Health care use
Felder et al [30] 2017	United States	• Web program	Mental health condition	Nonrandomized experimental	Women	27	• Medical • Mental health

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Ferrara et al [31] 2012	United States	• Telehealth	GDM	RCT	Women	1000	• Medical • Number of readings
Ferrara et al [32] 2016	United States	• Telehealth	GDM	RCT	Women	523	• Medical • Health behavior
Forsell et al [33] 2017	Sweden	• Telehealth • Web program	Mental health condition	RCT	Women	72	• Patient-centered outcomes • Mental health
Garnweidner-Holme et al [34] 2020	Norway	• Remote monitoring • Smartphone/tablet application	GDM	RCT	Women	238	• Health behavior
Ghaderi et al [35] 2019	Iran	• Smartphone/tablet application	GDM	RCT	Women	134	• Health knowledge/attitudes
Goetz et al [36] 2020	Taiwan	• Smartphone/tablet application	Mental health condition	Cohort	Women	30	• Mental health
Gong et al [37] 2021	China	• SMS text messaging	Mental health condition	RCT	Women	291	• Medical • Mental health
Guille et al [38] 2021	United States	• SMS text messaging • Telehealth	Mental health condition	Nonrandomized experimental	Women	2988	• Quality of care • Health behavior • Health care use
Guo et al [39] 2019	China	• Smartphone/tablet application	GDM	RCT	Women	172	• Medical • Number of readings • Health behavior • Health care use
Hantsoo et al [40] 2018	United States	• Smartphone/tablet application	Mental health condition	RCT	Women	61	• Health behavior • Health care use • Quality of care
Hedderson et al [41] 2018	United States	• DSS	GDM	RCT	Women	2014	• Medical
Heller et al [42] 2020	Netherlands	• Web program	Mental health condition	RCT	Women	159	• Mental health
Homko et al [43] 2007	United States	• Web program	GDM	RCT	Women	57	• Medical • Number of readings • Health knowledge/attitudes
Homko et al [44] 2012	United States	• Telehealth	GDM	RCT	Women	253	• Medical • Number of readings
Hoppe et al [45] 2020	United States	• Telehealth	HDP	Nonrandomized experimental	Women	428	• Medical • Number of readings • Health care use
	China		GDM	RCT	Women	309	• Medical

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Huang et al [46] 2021		• Smartphone/tablet application					
Jannati et al [47] 2020	Iran	• Smartphone/tablet application	Mental health condition	RCT	Women	75	• Mental health
Klokkenga et al [48] 2019	Ghana	• Smartphone/tablet application	Hemorrhage	RCT	Women	146	• Medical
Krishnamurti et al [49] 2021	United States	• Smartphone/tablet application	HDP	Cohort	Women	2567	• Health behavior • Quality of care
Lanssens et al [50] 2017	Belgium	• Remote monitoring	HDP	Retrospective	Women	166	• Medical • Health care use • Quality of care
Lanssens et al [51] 2018	Belgium	• Remote monitoring	HDP	Retrospective	Women	320	• Medical • Health care use • Quality of care
Latendresse et al [52] 2021	United States	• Telehealth	Mental health condition	Nonrandomized experimental	Women	47	• Mental health
Lewey et al [53] 2022	United States	• SMS text messaging • Web program • Wearable	HDP	RCT	Women	127	• Medical • Health behavior
Li et al [54] 2021	China	• Telehealth	GDM	RCT	Women	287	• Medical • Health behavior
Lim et al [55] 2021	Singapore	• Smartphone/tablet application	GDM	RCT	Women	200	• Medical • Mental health
Mackillop et al [56] 2018	United Kingdom	• Telehealth • SMS text messaging	GDM	RCT	Women	203	• Medical • Number of readings • Patient-centered outcomes
Milgrom et al [57] 2016	Australia	• Web program	Mental health condition	RCT	Women	43	• Mental health
Milgrom et al [58] 2021	Australia	• Telehealth	Mental health condition	RCT	Women	116	• Mental health
Miremberg et al [59] 2018	Israel	• Smartphone/tablet application	GDM	RCT	Women	120	• Medical • Number of readings • Patient-centered outcomes • Health behavior
Missler et al [60] 2020	Netherlands	• Telehealth	Mental health condition	RCT	Women	130	• Mental health • Parenting measures
Ngai et al [61] 2015	China	• Telehealth	Mental health condition	RCT	Women	397	• Mental health

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Nieminen et al [62] 2016	Sweden	• Web program	Mental health condition	RCT	Women	43	• Patient-centered outcomes • Mental health • Health behavior
Niksalehi et al [63] 2018	Iran	• SMS text messaging	Mental health condition	Nonrandomized experimental	Women	54	• Mental health
Nishimwe et al [64] 2021	Rwanda	• Smartphone/tablet application	Hemorrhage	Nonrandomized experimental	Providers	54	• Health knowledge/attitudes • Quality of care
Parsa et al [65] 2019	Iran	• Smartphone/tablet application	HDP	Nonrandomized experimental	Women	110	• Health knowledge/attitudes
Perry et al [66] 2018	United Kingdom	• Remote monitoring • Smartphone/tablet application	HDP	Case control	Women	108	• Medical • Number of readings • Health care use
Posmontier et al [67] 2016	United States	• Telehealth	Mental health condition	Cohort	Women	61	• Patient-centered outcomes • Mental health • Quality of care
Potzel et al [68] 2022	Germany	• Smartphone/tablet application	GDM	RCT	Women	203	• Medical • Number of readings • Health behavior
Rahman et al [69] 2017	Bangladesh	• Web program	Hemorrhage	RCT	Women	310	• Quality of care
Rasekaba et al [70] 2018	Australia	• Telehealth	GDM	RCT	Women	95	• Medical • Health care use
Sawyer et al [71] 2019	Australia	• Smartphone/tablet application	Mental health condition	RCT	Women	133	• Mental health
Shamshiri Milani et al [72] 2015	Iran	• Telehealth	Mental health condition	RCT	Women	126	• Mental health
Silva-Jose et al [73] 2021	Spain	• Telehealth • Web program	HDP	RCT	Women	206	• Medical • Health behavior
Simhi et al [74] 2021	Israel	• Telehealth	Mental health condition	RCT	Women	77	• Mental health • Health behavior
Skar et al [75] 2018	Norway	• Smartphone/tablet application	GDM	Qualitative	Women	17	• Mental health • Health knowledge/attitudes
St Pierre et al [76] 2017	Germany	• Anesthesia machine cognitive aid	Myocardial infarction	RCT	Women	83	• Quality of care
Sun et al [77] 2021	China	• Smartphone/tablet application	Mental health condition	RCT	Women	186	• Mental health
	Korea		GDM	RCT	Women	21	• Medical

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Sung et al [78] 2019		• Remote monitoring • Smartphone/tablet application					
Sung et al [79] 2021	Korea	• Telehealth	GDM	Cohort	Women	176	• Medical • Health care use
Tian et al [80] 2020	China	• Smartphone/tablet application	GDM	RCT	Women	169	• Health behavior
Tian et al [81] 2021	China	• Telehealth	GDM	RCT	Women	309	• Medical
Triebwasser et al [82] 2020	United States	• SMS text messaging • Web-based program	HDP	RCT	Women	333	• Number of readings
Tucker et al [83] 2022	United Kingdom	• Remote monitoring • Smartphone/tablet application	HDP	RCT	Women	2346	• Medical • Mental health • Quality of care
Ugarriza and Schmidt [84] 2006	United States	• Telehealth	Mental health condition	RCT	Women	30	• Mental health
Vandenberk et al [85] 2019	Belgium	• Remote monitoring	HDP	RCT	Women	108	• Number of readings • Mental health
van den Heuvel et al [86] 2020	Netherlands	• Remote monitoring • Smartphone/tablet application	HDP	Case control	Women	133	• Medical • Health care use
van Montfort et al [87] 2020	Netherlands	• Web program	HDP	Cohort	Women	850	• Quality of care
Van Ryswyk et al [88] 2015	New Zealand	• SMS text messaging	GDM	RCT	Women	268	• Medical • Health behavior
Vigod et al [89] 2021	Canada	• Web program	Mental health condition	RCT	Women	98	• Mental health
Wozney et al [90] 2017	Canada	• Telehealth	Mental health condition	RCT	Women	62	• Mental health
Yang et al [91] 2018	China	• Smartphone/tablet application	GDM	RCT	Women	107	• Medical
Yew et al [92] 2021	Singapore	• Smartphone/tablet application	GDM	RCT	Women	340	• Medical
Zera et al [93] 2015	United States	• DSS	GDM	RCT	Women	847	• Quality of care
Zhuo et al [94] 2022	China	• Smartphone/tablet application	GDM	RCT	Women	124	• Medical • Health behavior

[a]GDM: gestational diabetes mellitus.

[b]RCT: randomized controlled trial.

[c]DSS: decision support system.

[d]HDP: hypertensive disorders of pregnancy.

[e]PTB: preterm birth.

Table 2. Outcomes measured.

	Frequency, n (%)
Study design (n=80)	
Randomized controlled trial	60 (75)
Nonrandomized experimental	9 (11)
Cohort	6 (8)
Case control	2 (3)
Retrospective	2 (3)
Qualitative	1 (1)
Region (n=80)	
North America	24 (30)
East Asia and Pacific	22 (28)
Europe and Central Asia	20 (25)
Middle East and North Africa	9 (11)
Sub-Saharan Africa	3 (4)
South Asia	2 (3)
Latin America and Caribbean	0 (0)
Technology (n=87)	
Application	28 (32)
Telehealth/telemedicine	23 (26)
Remote monitoring	12 (13)
Web-based program	10 (11)
SMS text messaging	8 (9)
EHR[a] DSS[b]	3 (3)
Wearable	2 (2)
Other	1 (1)
Perinatal health outcome (n=82)	
New mental health diagnosis	30 (37)
GDM[c]	30 (37)
Preeclampsia/HDP[d]	12 (15)
Hypertension	5 (6)
Hemorrhage	3 (4)
Preterm birth	1 (1)
Myocardial infarction	1 (1)
Study population (n=80)	
Women	77 (96)
Providers	3 (4)
Outcomes (n=140)	
Medical outcomes	39 (27.9)
Mental health outcomes	30 (21.4)
Health behavior	21 (15.0)
Quality of care	19 (13.6)
Number of readings	13 (9.3)

	Frequency, n (%)
Health care use	12 (8.6)
Health knowledge/attitudes	5 (3.6)
Other	1 (0.7)
Results (n=80)	
Significant for at least one measure	63 (79)
No significance	17 (21)

[a]EHR: electronic health record.

[b]DSS: decision support system.

[c]GDM: gestational diabetes mellitus.

[d]HDP: hypertensive disorders of pregnancy.

Included studies were mostly RCTs (n=60). Other study designs included nonrandomized experimental (n=9), cohort (n=6), case control (n=2), retrospective (n=2), and qualitative (n=1). Included studies represented a range of countries and regions. Most of the studies were conducted in North America (n=24), East Asia and the Pacific (n=22), and Western Europe (n=20). Fewer studies were authored from the Middle East and North Africa (n=9), sub-Saharan Africa (n=3), and South Asia (n=2). No studies focused on Latin America or the Caribbean.

Informatics Technologies

The specific informatics technologies tested by the study were described. Descriptive categories included telehealth or telemedicine (remote consultation via telephone or video chat that may replace medical consultation), remote monitoring (patient-performed measurements of vital signs and blood glucose), electronic health record decision support system (tools within the electronic health record that assist providers with clinical decisions), smartphone or tablet applications, SMS text messages (between providers and patients or between patients), web-based programs (patient portals and other programs administered through a web browser, not through a smartphone or tablet application), and other (with the technology charted manually). Studies tested a range of informatics technologies. The most frequently tested technologies included smartphone/tablet applications (n=28) and telehealth/telemedicine (n=23). Technologies tested less frequently included remote monitoring (n=15), web-based programs (n=10), and SMS text messaging (n=8). Technologies assessed in very few studies included electronic health record decision support (n=3), wearable technologies (n=2), and other (cognitive aid on an anesthesia machine; n=1).

Health Problems

Articles were categorized by the health problems treated by the informatics technology; some tested more than one. These included preeclampsia or hypertensive disorders of pregnancy, hypertension, GDM, PTB, hemorrhage, and new (perinatal) mental health diagnoses. Although our protocol searched for "severe maternal morbidity as defined by the CDC" and included a list of 21 conditions [13], after reviewing the articles, it was clear that the only condition from the CDC list that was frequently tested by informatics technology was postpartum hemorrhage. Thus, any additional conditions were added manually in the "other" option. Studies primarily used informatics technologies to address new mental health conditions (n=30) and GDM (n=30). Other health problems addressed by the studies include preeclampsia or hypertensive disorders of pregnancy (n=12), hypertension (n=5), and hemorrhage (n=3). One study each addressed preterm birth and myocardial infarction.

Population Type

Population type was categorized as either: women or medical providers who treat women. Included studies overwhelmingly tested informatics technologies on women, with only 3 studies testing informatics technologies on providers. US studies were also coded for sample demographics.

Outcomes

Finally, outcomes were categorized to gain an understanding of the measures upon which the studies focused; many tested more than one measure. These included medical outcomes (eg, weight loss, blood pressure, blood glucose), number of remote measures (eg, patient-collected and -reported blood pressure or blood glucose measurements), mental health outcomes (eg, Edinburgh Postnatal Depression Scale score, subjective well-being measures), health behaviors (eg, dietary habits, exercise frequency), health knowledge or attitudes (eg, patient's understanding of health condition, patient-reported attitude toward care), health care use (eg, number of emergency room visits, number of prenatal care visits), quality of care (eg, Hospital Consumer Assessment of Healthcare Providers and Systems [HCAHPS] scores, safety of care, effectiveness of care, efficiency of care), or other. Studies measured myriad outcomes, and many studies measured more than one outcome. Many assessed medical outcomes (n=39), mental health outcomes (n=30), and patient health behaviors (n=21). Fewer measured quality of care (n=19), number of remote readings (n=13), and health care use (n=12). Very few measured health knowledge and attitudes (n=5) or other outcomes (n=1).

Effect

Studies were coded as to whether a significant positive effect was exhibited for at least one health outcome measure (yes/no). Most (n=63) studies found that the informatics technology yielded at least one positive effect for a health-related (ie, not user experience) outcome. Some (n=17) studies showed no significant positive effect of the informatics technology tested.

Discussion

Principal Findings

This scoping review was performed to characterize what informatics interventions have been tested to address the maternal morbidities that result in long-term negative health outcomes for mothers. Our results show that most of the tested informatics technologies were tested in North America, East Asia, and Western Europe. They overwhelmingly focused on patient populations diagnosed with GDM, mental health conditions, and preeclampsia. Most were patient focused and most aimed to prevent, diagnose, or treat physical or mental health outcomes.

Women have been entering pregnancy less healthy than previous generations. As the world population grows increasingly obese, high BMI presents critical risks for birthing people, with obesity linked to such perinatal complications as gestational diabetes [95], preeclampsia [96], and postpartum hemorrhage [97]. Women also currently enter pregnancy in poorer cardiovascular health than previous generations, with prepregnancy hypertension especially pronounced in rural populations of childbearing age [98]. This scoping review surveys the landscape of informatics interventions tested to treat these increasingly common conditions.

Most of the technologies were tested in North America, East Asia or the Pacific, and Western Europe. Despite some evidence of resistance to new technologies [99], studies generally show perceptions of mobile health to be positive in sub-Saharan Africa, especially toward maternal health applications [100]. There may be opportunities to test technologies developed in North America, Asia, and Europe in underserved parts of the world. In addition to improving maternal health in these underserved areas, testing in diverse types of populations may yield compelling findings that could be translated into US, Asian, and European contexts through a process of reciprocal innovation [101].

Our results show that most of the tested informatics technologies focused on patient populations diagnosed with GDM; mental health conditions; and, to a lesser extent, preeclampsia. Far fewer focused on preventing or treating postpartum hemorrhage. Although prevalences of GDM and mental health conditions are high, preeclampsia and hemorrhage are leading causes of maternal mortality and pregnancy-associated deaths in both the United States and abroad [101]. Postpartum hemorrhage, in particular, is the leading cause of maternal death worldwide, causing 94% of maternal deaths [102,103].

There are at least two possible reasons for this imbalance in testing of interventions. First, diabetes mellitus and mental health conditions are also common in the general population, unlike preeclampsia and hemorrhage. Thus, preexisting interventions may be more easily transferrable to pregnant populations. Second, GDM and postpartum depression can be diagnosed through routine screening, with GDM diagnosed via prenatal glucose testing and postpartum depression diagnosed using the Edinburgh Postnatal Depression Scale. Indeed, 9.3% of interventions measured the extent/accuracy with which patients review their own biodata, including blood glucose levels. Screening for GDM could be seen as a parsimonious method to improve women's health. On the other hand, postpartum hemorrhage is less predictable, typically being diagnosed through an urgent/emergent presentation. Thus, the ease of routine screening practices may lend certain diagnoses to being more amenable to the implementation of informatics interventions.

Most of the interventions in this review aimed to prevent, diagnose, or treat physical or mental health outcomes, and almost all were patient focused. Very few interventions (n=3, 4%) were tested on providers who treat women. Of those studies, two of the interventions focused on midwives and one used a decision support system to assist physicians in screening for maternal depression. Most interventions in this review were tested on women and involved active participation from patients (eg, taking blood glucose readings, logging symptoms in an app, participating in an online support group). In other words, most of the interventions entrusted patients with aspects of their own health.

This focus on outcomes is logical given the maternal mortality crisis both within and outside of the United States. Mental health conditions are the leading cause of pregnancy-related deaths in the United States according to state Maternal Mortality Review Committees [104] and of non-Hispanic White and Hispanic women when stratified by ethnicity. Cardiac conditions are the leading cause of pregnancy-related death for Black/African American women in the United States, but the focus on interventions aimed at patients themselves minimizes the role provider knowledge and practice within a health system plays in the crisis. Research shows that lack of standardized emergency obstetric care, maternity care coordination [105], and systemic racism [106] can play a role in the safe and effective delivery of care. The results of this scoping review show a large gap in the literature related to using informatics to improve the quality of care for maternal morbidities. Interventions such as telehealth have been shown to improve obstetric quality, including perinatal smoking cessation and breastfeeding [107]. It is possible that expanding such technologies and examining the way they are used by physicians can help improve the quality of care and ultimately save the lives of women.

The need for informatics interventions does not end with birth. Prior research suggests that many physicians, including obstetricians, are unaware of how maternal morbidity can influence maternal health in the decades following pregnancy [108]. Because of this lack of awareness, patients often do not know to monitor their own health in the long term, and they may not receive referrals for proper follow-up services. The paucity of provider-focused technologies identified in this review may also present opportunities for the development of informatics-based interventions targeting physicians and affiliated providers—especially in the treatment and prevention of postpartum preeclampsia and hemorrhage—that could use technologies like decision support to improve maternal health.

Limitations

Our study has several limitations. Because this is a scoping review, we did not systematically assess the quality of individual studies. The studies reviewed may vary in quality or contain potential biases or methodological limitations. Therefore, this study is not intended to assess the quality of the evidence supporting informatics-based interventions on the given perinatal health conditions. Additionally, the nature of a scoping review is not to analytically aggregate studies so as to make claims about effect sizes. When additional, more comparable studies have accumulated, further research is needed to examine which informatics interventions are most effective and the size of the impact on maternal and perinatal health.

In addition, we only considered published studies that tested an informatics intervention and presented health-related outcomes. We excluded many studies because they tested feasibility or acceptability. It is possible that such excluded feasibility studies will eventually test health-related outcomes. Overall, this scoping review summarizes trends in the populations, geographic areas, technologies, and conditions targeted by informatics interventions tested and disseminated in the available literature.

Conclusions

This scoping review paints a picture of the landscape of informatics interventions aimed at preventing and treating maternal morbidity. Most interventions identified in this study were tested in North America, East Asia and the Pacific, or Western Europe. Most tested either smartphone/tablet applications or telehealth/telemedicine, and most technologies tested for new mental health conditions and GDM. Almost all the studies tested technologies on populations of women and reported medical, mental health, and patient behavior outcomes. Results suggest that there may be opportunities to use informatics technologies to target providers who treat women as well as conditions such as postpartum hemorrhage that are more likely to lead to mortality. As the landscape of informatics applications in health care continues to expand, maternal health is poised to be an important target of these interventions.

Acknowledgments

The authors would like to acknowledge the VA Medical Informatics Fellowship. This work was funded by the Department of Veterans Affairs Office of Academic Affiliations. The views expressed in this article are those of the authors and do not reflect the position or policy of the US Department of Veterans Affairs, the US government, or the authors' affiliated institutions.

Data Availability

The search strategies used in this study are openly available at searchRxiv [11,109-111].

Authors' Contributions

Conceptualization: JI (lead), DAH (equal), JCS (supporting)
Formal analysis: JI (lead), RCS (supporting), SS (supporting), DAH (supporting)
Supervision: DAH (lead), JI (equal)
Writing – original draft: JI (lead), JCS (supporting)
Writing – review and editing: JI (lead), JCS (supporting), DAH (supporting)
Methodology: JCS (lead), DAH (supporting)
Investigation: RCS (lead), SS (equal), SAEA (supporting), OO (supporting), BB (supporting)
Generative artificial intelligence was not used in the writing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 26 KB - ijmr_v14i1e64826_app1.docx]

References

1. Chivers BR, Boyle JA, Lang AY, Teede HJ, Moran LJ, Harrison CL. Preconception health and lifestyle behaviours of women planning a pregnancy: a cross-sectional study. J Clin Med 2020 Jun 02;9(6):1701 [FREE Full text] [doi: 10.3390/jcm9061701] [Medline: 32498329]
2. Bellamy L, Casas J, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. BMJ 2007 Nov 10;335(7627):974 [FREE Full text] [doi: 10.1136/bmj.39335.385301.BE] [Medline: 17975258]
3. Neiger R. Long-term effects of pregnancy complications on maternal health: a review. J Clin Med 2017 Jul 27;6(8):76. [doi: 10.3390/jcm6080076] [Medline: 28749442]

4. Ukah UV, Dayan N, Potter BJ, Ayoub A, Auger N. Severe maternal morbidity and risk of mortality beyond the postpartum period. Obstet Gynecol 2021 Feb 01;137(2):277-284. [doi: 10.1097/AOG.0000000000004223] [Medline: 33416296]
5. Robbins CL, Hutchings Y, Dietz PM, Kuklina EV, Callaghan WM. History of preterm birth and subsequent cardiovascular disease: a systematic review. Am J Obstet Gynecol 2014 Apr;210(4):285-297 [FREE Full text] [doi: 10.1016/j.ajog.2013.09.020] [Medline: 24055578]
6. Iftikhar P, Kuijpers MV, Khayyat A, Iftikhar A, DeGouvia De Sa M. Artificial intelligence: a new paradigm in obstetrics and gynecology research and clinical practice. Cureus 2020 Feb 28;12(2):e7124 [FREE Full text] [doi: 10.7759/cureus.7124] [Medline: 32257670]
7. Reed T. Health gaps emerge in how doctors and patients see AI. Axios. 2024 Apr 15. URL: https://www.axios.com/2024/04/15/doctors-patients-ai [accessed 2024-04-18]
8. Jean-Francois B, Bailey Lash T, Dagher RK, Green Parker MC, Han SB, Lewis Johnson T. The potential for health information technology tools to reduce racial disparities in maternal morbidity and mortality. J Womens Health (Larchmt) 2021 Feb;30(2):274-279 [FREE Full text] [doi: 10.1089/jwh.2020.8889] [Medline: 33211604]
9. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
10. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol 2009 Sep;62(9):944-952. [doi: 10.1016/j.jclinepi.2008.10.012] [Medline: 19230612]
11. Stumpff JC, Inderstrodt J, Haggstrom D. Systematic search: severe maternal morbidity and informatics (Engineering Village). searchRxiv Preprint posted online on May 11, 2023. [doi: 10.1079/searchrxiv.2023.00212]
12. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Implement 2021 Mar;19(1):3-10. [doi: 10.1097/XEB.0000000000000277] [Medline: 33570328]
13. Inderstrodt J. Maternal morbidity and informatics scoping review. Open Science Framework. 2024 Dec 20. URL: https://osf.io/nfvce [accessed 2025-02-13]
14. Identifying severe maternal morbidity (SMM). Centers for Disease Control and Prevention. 2024 May 15. URL: https://www.cdc.gov/maternal-infant-health/php/severe-maternal-morbidity/icd.html [accessed 2024-06-10]
15. Abbaspoor Z, Amani A, Afshari P, Jafarirad S. The effect of education through mobile phone short message service on promoting self-care in pre-diabetic pregnant women: a randomized controlled trial. J Telemed Telecare 2020 May;26(4):200-206. [doi: 10.1177/1357633X18791419] [Medline: 30193565]
16. Abejirinde IO, De Brouwere V, van Roosmalen J, van der Heiden M, Apentibadek N, Bardají A, et al. Viability of diagnostic decision support for antenatal care in rural settings: findings from the Bliss4Midwives Intervention in Northern Ghana. J Glob Health 2019 Jun;9(1):010420 [FREE Full text] [doi: 10.7189/jogh.09.010420] [Medline: 30937164]
17. Al-Ofi EA, Mosli HH, Ghamri KA, Ghazali SM. Management of postprandial hyperglycaemia and weight gain in women with gestational diabetes mellitus using a novel telemonitoring system. J Int Med Res 2019 Feb;47(2):754-764 [FREE Full text] [doi: 10.1177/0300060518809872] [Medline: 30442052]
18. Arias MP, Wang E, Leitner K, Sannah T, Keegan M, Delferro J, et al. The impact on postpartum care by telehealth: a retrospective cohort study. Am J Obstet Gynecol MFM 2022 May;4(3):100611 [FREE Full text] [doi: 10.1016/j.ajogmf.2022.100611] [Medline: 35331971]
19. Bartholomew ML, Soules K, Church K, Shaha S, Burlingame J, Graham G, et al. Managing diabetes in pregnancy using cell phone/internet technology. Clin Diabetes 2015 Oct;33(4):169-174 [FREE Full text] [doi: 10.2337/diaclin.33.4.169] [Medline: 26487790]
20. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital peer-support platform (7Cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
21. Bellad MB, Goudar SS, Mallapur AA, Sharma S, Bone J, Charantimath US, CLIP India Working Group (Table S1). Community level interventions for pre-eclampsia (CLIP) in India: a cluster randomised controlled trial. Pregnancy Hypertens 2020 Jul;21:166-175 [FREE Full text] [doi: 10.1016/j.preghy.2020.05.008] [Medline: 32554291]
22. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
23. Carlisle N, Watson HA, Seed PT, Carter J, Kuhrt K, Tribe RM, et al. Impact of a medical mobile phone app (QUiPP) for predicting preterm birth on the anxiety and decisional conflicts faced by women in threatened preterm labour. Midwifery 2021 Jan;92:102864. [doi: 10.1016/j.midw.2020.102864] [Medline: 33137547]
24. Carroll AE, Biondich P, Anand V, Dugan TM, Downs SM. A randomized controlled trial of screening for maternal depression with a clinical decision support system. J Am Med Inform Assoc 2013;20(2):311-316 [FREE Full text] [doi: 10.1136/amiajnl-2011-000682] [Medline: 22744960]

25. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]
26. Chappell L, Tucker K, Galal U, Yu LM, Campbell H, Rivero-Arias O, et al. Effect of self-monitoring of blood pressure on blood pressure control in pregnant individuals with chronic or gestational hypertension: the BUMP 2 randomized clinical trial. JAMA 2022 May 03;327(17):1666-1678 [FREE Full text] [doi: 10.1001/jama.2022.4726] [Medline: 35503345]
27. Cheung NW, Blumenthal C, Smith BJ, Hogan R, Thiagalingam A, Redfern J, et al. A pilot randomised controlled trial of a text messaging intervention with customisation using linked data from wireless wearable activity monitors to improve risk factors following gestational diabetes. Nutrients 2019 Mar 11;11(3):590 [FREE Full text] [doi: 10.3390/nu11030590] [Medline: 30862052]
28. Dennis C, Hodnett E, Kenton L, Weston J, Zupancic J, Stewart DE, et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. BMJ 2009 Jan 15;338:a3064 [FREE Full text] [doi: 10.1136/bmj.a3064] [Medline: 19147637]
29. Dennis C, Grigoriadis S, Zupancic J, Kiss A, Ravitz P. Telephone-based nurse-delivered interpersonal psychotherapy for postpartum depression: nationwide randomised controlled trial. Br J Psychiatry 2020 Apr;216(4):189-196. [doi: 10.1192/bjp.2019.275] [Medline: 32029010]
30. Felder JN, Segal Z, Beck A, Sherwood NE, Goodman SH, Boggs J, et al. An open trial of web-based mindfulness-based cognitive therapy for perinatal women at risk for depressive relapse. Cognitive Behav Pract 2017 Feb;24(1):26-37. [doi: 10.1016/j.cbpra.2016.02.002]
31. Ferrara A, Hedderson MM, Ching J, Kim C, Peng T, Crites YM. Referral to telephonic nurse management improves outcomes in women with gestational diabetes. Am J Obstet Gynecol 2012 Jun;206(6):491.e1-491.e5 [FREE Full text] [doi: 10.1016/j.ajog.2012.04.019] [Medline: 22631866]
32. Ferrara A, Hedderson MM, Brown SD, Albright CL, Ehrlich SF, Tsai A, et al. The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster randomized controlled trial. Diabetes Care 2016 Jan;39(1):65-74 [FREE Full text] [doi: 10.2337/dc15-1254] [Medline: 26657945]
33. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
34. Garnweidner-Holme L, Henriksen L, Torheim LE, Lukasse M. Effect of the Pregnant+ smartphone app on the dietary behavior of women with gestational diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 04;8(11):e18614 [FREE Full text] [doi: 10.2196/18614] [Medline: 33146620]
35. Ghaderi M, Farahani MA, Hajiha N, Ghaffari F, Haghani H. The role of smartphone-based education on the risk perception of type 2 diabetes in women with gestational diabetes. Health Technol 2019 Aug 7;9(5):829-837. [doi: 10.1007/s12553-019-00342-3]
36. Goetz M, Schiele C, Müller M, Matthies LM, Deutsch TM, Spano C, et al. Effects of a brief electronic mindfulness-based intervention on relieving prenatal depression and anxiety in hospitalized high-risk pregnant women: exploratory pilot study. J Med Internet Res 2020 Aug 11;22(8):e17593 [FREE Full text] [doi: 10.2196/17593] [Medline: 32780023]
37. Gong M, Zhang S, Xi C, Luo M, Wang T, Wang Y, et al. Comprehensive intervention during pregnancy based on short message service to prevent or alleviate depression in pregnant women: a quasi-experimental study. Early Interv Psychiatry 2021 Apr;15(2):352-359. [doi: 10.1111/eip.12953] [Medline: 32232971]
38. Guille C, Maldonado L, Simpson AN, Newman R, King C, Cortese B, et al. A non-randomized trial of in-person versus text/telephone screening, brief intervention and referral to treatment for pregnant and postpartum women. Psychiatr Res Clin Pract 2021;3(4):172-183 [FREE Full text] [doi: 10.1176/appi.prcp.20210027] [Medline: 36101654]
39. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
40. Hantsoo L, Criniti S, Khan A, Moseley M, Kincler N, Faherty LJ, et al. A mobile application for monitoring and management of depressed mood in a vulnerable pregnant population. Psychiatr Serv 2018 Jan 01;69(1):104-107 [FREE Full text] [doi: 10.1176/appi.ps.201600582] [Medline: 29032705]
41. Hedderson MM, Brown SD, Ehrlich SF, Tsai A, Zhu Y, Quesenberry CP, et al. A tailored letter based on electronic health record data improves gestational weight gain among women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster-randomized controlled trial. Diabetes Care 2018 Jul;41(7):1370-1377 [FREE Full text] [doi: 10.2337/dc17-1133] [Medline: 29669736]
42. Heller HM, Hoogendoorn AW, Honig A, Broekman BFP, van Straten A. The effectiveness of a guided internet-based tool for the treatment of depression and anxiety in pregnancy (MamaKits Online): randomized controlled trial. J Med Internet Res 2020 Mar 23;22(3):e15172 [FREE Full text] [doi: 10.2196/15172] [Medline: 32202505]

43. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
44. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
45. Hoppe KK, Thomas N, Zernick M, Zella JB, Havighurst T, Kim K, et al. Telehealth with remote blood pressure monitoring compared with standard care for postpartum hypertension. Am J Obstet Gynecol 2020 Oct;223(4):585-588. [doi: 10.1016/j.ajog.2020.05.027] [Medline: 32439388]
46. Huang F, Zhang S, Tian Y, Li L, Li Y, Chen X, et al. Effect of mobile health based peripartum management of gestational diabetes mellitus on postpartum diabetes: a randomized controlled trial. Diabetes Res Clin Pract 2021 May;175:108775 [FREE Full text] [doi: 10.1016/j.diabres.2021.108775] [Medline: 33771645]
47. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: a randomized controlled trial. Int J Med Inform 2020 Sep;141:104145. [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
48. Klokkenga C, Enemark U, Adanu R, Lund S, Sørensen B, Attermann J. The effect of smartphone training of Ghanaian midwives by the Safe Delivery application on the incidence of postpartum hemorrhage: a cluster randomised controlled trial. Cogent Med 2019 Jul 17;6(1):1632016. [doi: 10.1080/2331205x.2019.1632016]
49. Krishnamurti T, Davis AL, Rodriguez S, Hayani L, Bernard M, Simhan HN. Use of a smartphone app to explore potential underuse of prophylactic aspirin for preeclampsia. JAMA Netw Open 2021 Oct 01;4(10):e2130804 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.30804] [Medline: 34714341]
50. Lanssens D, Vandenberk T, Smeets CJ, De Cannière H, Molenberghs G, Van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e25 [FREE Full text] [doi: 10.2196/mhealth.6552] [Medline: 28279948]
51. Lanssens D, Vonck S, Storms V, Thijs IM, Grieten L, Gyselaers W. The impact of a remote monitoring program on the prenatal follow-up of women with gestational hypertensive disorders. Eur J Obstet Gynecol Reprod Biol 2018 Apr;223:72-78 [FREE Full text] [doi: 10.1016/j.ejogrb.2018.02.015] [Medline: 29500948]
52. Latendresse G, Bailey E, Iacob E, Murphy H, Pentecost R, Thompson N, et al. A group videoconference intervention for reducing perinatal depressive symptoms: a telehealth pilot study. J Midwifery Womens Health 2021 Jan;66(1):70-77 [FREE Full text] [doi: 10.1111/jmwh.13209] [Medline: 33576146]
53. Lewey J, Murphy S, Zhang D, Putt ME, Elovitz MA, Riis V, et al. Effectiveness of a text-based gamification intervention to improve physical activity among postpartum individuals with hypertensive disorders of pregnancy: a randomized clinical trial. JAMA Cardiol 2022 Jun 01;7(6):591-599 [FREE Full text] [doi: 10.1001/jamacardio.2022.0553] [Medline: 35442393]
54. Li M, Lin Q, Shi J, Xi Y, Xiang C, Yong C, et al. The impact of lifestyle intervention on dietary quality among rural women with previous gestational diabetes mellitus-a randomized controlled study. Nutrients 2021 Jul 30;13(8):2642 [FREE Full text] [doi: 10.3390/nu13082642] [Medline: 34444802]
55. Lim K, Chan S, Lim SL, Tai BC, Tsai C, Wong SR, et al. A Smartphone App to Restore Optimal Weight (SPAROW) in women with recent gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 Mar 16;9(3):e22147 [FREE Full text] [doi: 10.2196/22147] [Medline: 33724204]
56. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
57. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 07;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
58. Milgrom J, Danaher BG, Seeley JR, Holt CJ, Holt C, Ericksen J, et al. Internet and face-to-face cognitive behavioral therapy for postnatal depression compared with treatment as usual: randomized controlled trial of MumMoodBooster. J Med Internet Res 2021 Dec 08;23(12):e17185 [FREE Full text] [doi: 10.2196/17185] [Medline: 34889742]
59. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
60. Missler M, van Straten A, Denissen J, Donker T, Beijers R. Effectiveness of a psycho-educational intervention for expecting parents to prevent postpartum parenting stress, depression and anxiety: a randomized controlled trial. BMC Pregnancy Childbirth 2020 Oct 31;20(1):658 [FREE Full text] [doi: 10.1186/s12884-020-03341-9] [Medline: 33129314]
61. Ngai F, Wong PW, Leung K, Chau P, Chung K. The effect of telephone-based cognitive-behavioral therapy on postnatal depression: a randomized controlled trial. Psychother Psychosom 2015;84(5):294-303. [doi: 10.1159/000430449] [Medline: 26278623]

62. Nieminen K, Berg I, Frankenstein K, Viita L, Larsson K, Persson U, et al. Internet-provided cognitive behaviour therapy of posttraumatic stress symptoms following childbirth-a randomized controlled trial. Cogn Behav Ther 2016 Jun;45(4):287-306. [doi: 10.1080/16506073.2016.1169626] [Medline: 27152849]
63. Niksalehi S, Taghadosi M, Mazhariazad F, Tashk M. The effectiveness of mobile phone text massaging support for mothers with postpartum depression: a clinical before and after study. J Family Med Prim Care 2018;7(5):1058-1062 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_120_17] [Medline: 30598957]
64. Nishimwe A, Ibisomi L, Nyssen M, Conco DN. The effect of an mLearning application on nurses' and midwives' knowledge and skills for the management of postpartum hemorrhage and neonatal resuscitation: pre-post intervention study. Hum Resour Health 2021 Jan 26;19(1):14 [FREE Full text] [doi: 10.1186/s12960-021-00559-2] [Medline: 33499870]
65. Parsa S, Khajouei R, Baneshi MR, Aali BS. Improving the knowledge of pregnant women using a pre-eclampsia app: a controlled before and after study. Int J Med Inform 2019 May;125:86-90. [doi: 10.1016/j.ijmedinf.2019.03.001] [Medline: 30914185]
66. Perry H, Sheehan E, Thilaganathan B, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population. Ultrasound Obstet Gynecol 2018 Apr;51(4):524-530 [FREE Full text] [doi: 10.1002/uog.19023] [Medline: 29468771]
67. Posmontier B, Neugebauer R, Stuart S, Chittams J, Shaughnessy R. Telephone-administered interpersonal psychotherapy by nurse-midwives for postpartum depression. J Midwifery Womens Health 2016 Jul;61(4):456-466. [doi: 10.1111/jmwh.12411] [Medline: 26970401]
68. Potzel AL, Gar C, Banning F, Sacco V, Fritsche A, Fritsche L, et al. A novel smartphone app to change risk behaviors of women after gestational diabetes: a randomized controlled trial. PLoS One 2022;17(4):e0267258 [FREE Full text] [doi: 10.1371/journal.pone.0267258] [Medline: 35476681]
69. Rahman A, Akhter S, Nisha MK, Islam SS, Ashraf F, Rahman M, et al. Can mHealth improve access to safe blood for transfusion during obstetric emergency? Int J Womens Health 2017;9:235-243 [FREE Full text] [doi: 10.2147/IJWH.S120157] [Medline: 28461767]
70. Rasekaba TM, Furler J, Young D, Liew D, Gray K, Blackberry I, et al. Using technology to support care in gestational diabetes mellitus: quantitative outcomes of an exploratory randomised control trial of adjunct telemedicine for gestational diabetes mellitus (TeleGDM). Diabetes Res Clin Pract 2018 Aug;142:276-285. [doi: 10.1016/j.diabres.2018.05.049] [Medline: 29885390]
71. Sawyer A, Kaim A, Le H, McDonald D, Mittinty M, Lynch J, et al. The effectiveness of an app-based nurse-moderated program for new mothers with depression and parenting problems (eMums Plus): pragmatic randomized controlled trial. J Med Internet Res 2019 Jun 04;21(6):e13689 [FREE Full text] [doi: 10.2196/13689] [Medline: 31165715]
72. Shamshiri Milani H, Azargashb E, Beyraghi N, Defaie S, Asbaghi T. Effect of telephone-based support on postpartum depression: a randomized controlled trial. Int J Fertil Steril 2015;9(2):247-253 [FREE Full text] [doi: 10.22074/ijfs.2015.4246] [Medline: 26246884]
73. Silva-Jose C, Sánchez-Polán M, Diaz-Blanco Á, Coterón J, Barakat R, Refoyo I. Effectiveness of a virtual exercise program during COVID-19 confinement on blood pressure control in healthy pregnant women. Front Physiol 2021;12:645136 [FREE Full text] [doi: 10.3389/fphys.2021.645136] [Medline: 33776798]
74. Simhi M, Sarid O, Rowe H, Fisher J, Cwikel J. A cognitive-behavioral intervention for postpartum anxiety and depression: individual phone vs. group format. J Clin Med 2021 Dec 18;10(24):5952 [FREE Full text] [doi: 10.3390/jcm10245952] [Medline: 34945248]
75. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
76. St Pierre M, Luetcke B, Strembski D, Schmitt C, Breuer G. The effect of an electronic cognitive aid on the management of ST-elevation myocardial infarction during caesarean section: a prospective randomised simulation study. BMC Anesthesiol 2017 Mar 20;17(1):46 [FREE Full text] [doi: 10.1186/s12871-017-0340-4] [Medline: 28320312]
77. Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F. Effectiveness of smartphone-based mindfulness training on maternal perinatal depression: randomized controlled trial. J Med Internet Res 2021 Jan 27;23(1):e23410 [FREE Full text] [doi: 10.2196/23410] [Medline: 33502326]
78. Sung J, Lee DY, Min KP, Park C. Peripartum management of gestational diabetes using a digital health care service: a pilot, randomized controlled study. Clin Ther 2019 Nov;41(11):2426-2434. [doi: 10.1016/j.clinthera.2019.09.005] [Medline: 31587813]
79. Sung Y, Zhang D, Eswaran H, Lowery CL. Evaluation of a telemedicine program managing high-risk pregnant women with pre-existing diabetes in Arkansas's Medicaid program. Semin Perinatol 2021 Aug;45(5):151421. [doi: 10.1016/j.semperi.2021.151421] [Medline: 34274150]
80. Tian Y, Zhang S, Huang F, Shi F, Li Y, Chen X, et al. Glycemic qualification rate and frequency of self-monitoring blood glucose glycemic qualification rate and frequency of self-monitoring blood glucose (SMBG) in women with gestational diabetes mellitus (GDM). Diabetes Res Clin Pract 2020 Dec;170:108482. [doi: 10.1016/j.diabres.2020.108482] [Medline: 32998018]

81. Tian Y, Zhang S, Huang F, Ma L. Comparing the efficacies of telemedicine and standard prenatal care on blood glucose control in women with gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 25;9(5):e22881 [FREE Full text] [doi: 10.2196/22881] [Medline: 33783365]
82. Triebwasser JE, Janssen MK, Hirshberg A, Srinivas SK. Successful implementation of text-based blood pressure monitoring for postpartum hypertension. Pregnancy Hypertens 2020 Oct;22:156-159. [doi: 10.1016/j.preghy.2020.09.001] [Medline: 32980623]
83. Tucker KL, Mort S, Yu L, Campbell H, Rivero-Arias O, Wilson HM, et al. Effect of self-monitoring of blood pressure on diagnosis of hypertension during higher-risk pregnancy: the BUMP 1 randomized clinical trial. JAMA 2022 May 03;327(17):1656-1665 [FREE Full text] [doi: 10.1001/jama.2022.4712] [Medline: 35503346]
84. Ugarriza DN, Schmidt L. Telecare for women with postpartum depression. J Psychosoc Nurs Ment Health Serv 2006 Jan;44(1):37-45. [doi: 10.3928/02793695-20060101-08] [Medline: 16475443]
85. Vandenberk T, Lanssens D, Storms V, Thijs IM, Bamelis L, Grieten L, et al. Relationship between adherence to remote monitoring and patient characteristics: observational study in women with pregnancy-induced hypertension. JMIR Mhealth Uhealth 2019 Aug 28;7(8):e12574 [FREE Full text] [doi: 10.2196/12574] [Medline: 31464190]
86. van den Heuvel JFM, Lely AT, Huisman JJ, Trappenburg JCA, Franx A, Bekker MN. SAFE@HOME: digital health platform facilitating a new care path for women at increased risk of preeclampsia - a case-control study. Pregnancy Hypertens 2020 Oct;22:30-36 [FREE Full text] [doi: 10.1016/j.preghy.2020.07.006] [Medline: 32717653]
87. van Montfort P, Smits LJM, van Dooren IMA, Lemmens SMP, Zelis M, Zwaan IM, et al. Implementing a preeclampsia prediction model in obstetrics: cutoff determination and health care professionals' adherence. Med Decis Making 2020 Jan;40(1):81-89 [FREE Full text] [doi: 10.1177/0272989X19889890] [Medline: 31789093]
88. Van Ryswyk EM, Middleton PF, Hague WM, Crowther CA. Postpartum SMS reminders to women who have experienced gestational diabetes to test for type 2 diabetes: the DIAMIND randomized trial. Diabet Med 2015 Oct;32(10):1368-1376. [doi: 10.1111/dme.12769] [Medline: 25816702]
89. Vigod SN, Slyfield Cook G, Macdonald K, Hussain-Shamsy N, Brown HK, de Oliveira C, et al. Mother matters: pilot randomized wait-list controlled trial of an online therapist-facilitated discussion board and support group for postpartum depression symptoms. Depress Anxiety 2021 Aug;38(8):816-825 [FREE Full text] [doi: 10.1002/da.23163] [Medline: 33949762]
90. Wozney L, Olthuis J, Lingley-Pottie P, McGrath PJ, Chaplin W, Elgar F, et al. Strongest Families™ Managing Our Mood (MOM): a randomized controlled trial of a distance intervention for women with postpartum depression. Arch Womens Ment Health 2017 Aug;20(4):525-537. [doi: 10.1007/s00737-017-0732-y] [Medline: 28593360]
91. Yang P, Lo W, He Z, Xiao X. Medical nutrition treatment of women with gestational diabetes mellitus by a telemedicine system based on smartphones. J Obstet Gynaecol Res 2018 Jul;44(7):1228-1234. [doi: 10.1111/jog.13669] [Medline: 29797375]
92. Yew TW, Chi C, Chan S, van Dam RM, Whitton C, Lim CS, et al. A randomized controlled trial to evaluate the effects of a smartphone application-based lifestyle coaching program on gestational weight gain, glycemic control, and maternal and neonatal outcomes in women with gestational diabetes mellitus: the SMART-GDM study. Diabetes Care 2021 Feb;44(2):456-463 [FREE Full text] [doi: 10.2337/dc20-1216] [Medline: 33184151]
93. Zera CA, Bates DW, Stuebe AM, Ecker JL, Seely EW. Diabetes screening reminder for women with prior gestational diabetes: a randomized controlled trial. Obstet Gynecol 2015 Jul;126(1):109-114 [FREE Full text] [doi: 10.1097/AOG.0000000000000883] [Medline: 26241263]
94. Zhuo Y, Pan Y, Lin K, Yin G, Wu Y, Xu J, et al. Effectiveness of clinical pharmacist-led smartphone application on medication adherence, insulin injection technique and glycemic control for women with gestational diabetes receiving multiple daily insulin injection: a randomized clinical trial. Prim Care Diabetes 2022 Apr;16(2):264-270. [doi: 10.1016/j.pcd.2022.02.003] [Medline: 35168915]
95. Najafi F, Hasani J, Izadi N, Hashemi-Nazari S, Namvar Z, Mohammadi S, et al. The effect of prepregnancy body mass index on the risk of gestational diabetes mellitus: a systematic review and dose-response meta-analysis. Obes Rev 2019 Mar;20(3):472-486. [doi: 10.1111/obr.12803] [Medline: 30536891]
96. Poorolajal J, Jenabi E. The association between body mass index and preeclampsia: a meta-analysis. J Matern Fetal Neonatal Med 2016 Nov;29(22):3670-3676. [doi: 10.3109/14767058.2016.1140738] [Medline: 26762770]
97. Blomberg M. Maternal obesity and risk of postpartum hemorrhage. Obstet Gynecol 2011 Sep;118(3):561-568. [doi: 10.1097/AOG.0b013e31822a6c59] [Medline: 21860284]
98. Cameron NA, Molsberry R, Pierce JB, Perak AM, Grobman WA, Allen NB, et al. Pre-pregnancy hypertension among women in rural and urban areas of the United States. J Am Coll Cardiol 2020 Dec 01;76(22):2611-2619 [FREE Full text] [doi: 10.1016/j.jacc.2020.09.601] [Medline: 33183896]
99. Osei E, Mashamba-Thompson TP. Mobile health applications for disease screening and treatment support in low-and middle-income countries: a narrative review. Heliyon 2021 Mar;7(3):e06639 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06639] [Medline: 33869857]

100. Peprah P, Abalo EM, Agyemang-Duah W, Budu HI, Appiah-Brempong E, Morgan AK, et al. Lessening barriers to healthcare in rural Ghana: providers and users' perspectives on the role of mHealth technology. A qualitative exploration. BMC Med Inform Decis Mak 2020 Feb 10;20(1):27 [FREE Full text] [doi: 10.1186/s12911-020-1040-4] [Medline: 32041608]
101. Ruhl LJ, Kiplagat J, O'Brien R, Wools-Kaloustian K, Scanlon M, Plater D, et al. A Global Health Reciprocal Innovation grant programme: 5-year review with lessons learnt. BMJ Glob Health 2023 Nov;8(Suppl 7):e013585 [FREE Full text] [doi: 10.1136/bmjgh-2023-013585] [Medline: 37977589]
102. Trost S, Beauregard J, Chandra G, Njie F, Berry J, Harvey A, et al. Pregnancy-related deaths: data from Maternal Mortality Review Committees in 36 US states, 2017-2019. CHAMPS. URL: https://champsonline.org/assets/files/Resources/ClinicalDocs/DiseaseCondition/MaternalMortalityReviewCommitteePregnancyRelatedDeathsData2017-2019.pdf [accessed 2024-01-15]
103. James AH, Federspiel JJ, Ahmadzia HK. Disparities in obstetric hemorrhage outcomes. Res Pract Thromb Haemost 2022 Jan;6(1):e12656 [FREE Full text] [doi: 10.1002/rth2.12656] [Medline: 35146237]
104. Four in 5 pregnancy-related deaths in the U.S. are preventable. Centers for Disease Control and Prevention. 2022 Sep 19. URL: https://www.cdc.gov/media/releases/2022/p0919-pregnancy-related-deaths.html [accessed 2024-06-10]
105. Ahn R, Gonzalez GP, Anderson B, Vladutiu CJ, Fowler ER, Manning L. Initiatives to reduce maternal mortality and severe maternal morbidity in the United States : a narrative review. Ann Intern Med 2020 Dec 01;173(11 Suppl):S3-S10 [FREE Full text] [doi: 10.7326/M19-3258] [Medline: 33253021]
106. Hailu EM, Maddali SR, Snowden JM, Carmichael SL, Mujahid MS. Structural racism and adverse maternal health outcomes: a systematic review. Health Place 2022 Nov;78:102923 [FREE Full text] [doi: 10.1016/j.healthplace.2022.102923] [Medline: 36401939]
107. DeNicola N, Grossman D, Marko K, Sonalkar S, Butler Tobah YS, Ganju N, et al. Telehealth interventions to improve obstetric and gynecologic health outcomes: a systematic review. Obstet Gynecol 2020 Feb;135(2):371-382 [FREE Full text] [doi: 10.1097/AOG.0000000000003646] [Medline: 31977782]
108. Roth H, LeMarquand G, Henry A, Homer C. Assessing knowledge gaps of women and healthcare providers concerning cardiovascular risk after hypertensive disorders of pregnancy-a scoping review. Front Cardiovasc Med 2019;6:178 [FREE Full text] [doi: 10.3389/fcvm.2019.00178] [Medline: 31850374]
109. Stumpff JC. Systematic search: severe maternal morbidity and informatics (Cochrane CENTRAL). searchRxiv Preprint posted online on September 27, 2022. [doi: 10.1079/searchRxiv.2022.00024]
110. Stumpff JC. searchRxiv Preprint posted online on September 6, 2022. [doi: 10.1079/searchRxiv.2022.00017]
111. Stumpff JC, Inderstrodt JM, Haggstrom D. Systematic search: severe maternal morbidity and informatics (IEEE Xplore). searchRxiv Preprint posted online on May 11, 2023. [doi: 10.1079/searchRxiv.2023.00211]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
GDM: gestational diabetes mellitus
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
PCC: Population, Concept, and Context
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
PTB: preterm birth
SMM: severe maternal morbidity

Edited by T de Azevedo Cardoso; submitted 27.07.24; peer-reviewed by A Morgan, R Appiah-Kubi; comments to author 01.10.24; revised version received 20.12.24; accepted 21.01.25; published 25.03.25.

Please cite as:
Inderstrodt J, Stumpff JC, Smollen RC, Sridhar S, El-Azab SA, Ojo O, Bowns B, Haggstrom DA
Informatics Interventions for Maternal Morbidity: Scoping Review
Interact J Med Res 2025;14:e64826
URL: https://www.i-jmr.org/2025/1/e64826
doi:10.2196/64826
PMID:

Glucagon-Like Peptide-1 Receptor Agonists Combined With Personalized Digital Health Care for the Treatment of Metabolic Syndrome in Adults With Obesity: Retrospective Observational Study

Hala Zakaria, MSc; Hadoun Jabri, MD; Sheikha Alshehhi, BSc; Milena Caccelli, MD; Joelle Debs, BSc; Yousef Said, MD; Joudy Kattan, MSc, MD; Noah Almarzooqi; Ali Hashemi, MBA; Ihsan Almarzooqi, PhD

GluCare.Health, Jumeirah 1, Alwasl road, Dubai, UAE, Dubai, United Arab Emirates

Corresponding Author:
Hala Zakaria, MSc
GluCare.Health, Jumeirah 1, Alwasl road, Dubai, UAE, Dubai, United Arab Emirates

Abstract

Background: Metabolic syndrome (MetS) represents a complex and multifaceted health condition characterized by a clustering of interconnected metabolic abnormalities, including central obesity, insulin resistance, dyslipidemia, and hypertension. Effective management of MetS is crucial for reducing the risk of cardiovascular diseases and type 2 diabetes.

Objective: This study aimed to assess the effectiveness of combining glucagon-like peptide-1 (GLP-1) and dual gastric inhibitory polypeptide (GIP)/GLP-1 agonists with a continuous, digitally delivered behavioral change model by an integrated care team, in treating MetS among individuals with obesity.

Methods: The 6-month Zone.Health (meta[bolic]) weight loss program involved 51 participants (mean age 45, SD 10 years; mean BMI 35, SD 5 kg/m²), categorized by gender, and treated with either tirzepatide or semaglutide. Participants received continuous support via a digital health platform, which facilitated real time monitoring and personalized feedback from an integrated care team. Engagement levels with the digital platform, measured by the frequency of inbound interactions, were tracked and analyzed in relation to health outcomes.

Results: Tirzepatide reduced waist circumference (WC) by −18.08 cm, compared with −13.04 cm with semaglutide (P<.001). Triglycerides decreased significantly with both drugs, with tirzepatide showing a reduction of −64.42 mg/dL and semaglutide −70.70 mg/dL (P<.001). Tirzepatide generally showed more pronounced improvements in fasting glucose, blood pressure (BP), low-density lipoprotein, and total cholesterol compared with semaglutide. Higher engagement with the digital health platform showed significant difference among the 3 groups; the group with the highest level of app-based interactions (≥25 interactions) had the greatest WC reduction (mean −19.04, SD 7.40 cm) compared with those with ≤15 interactions (mean −9.60, SD 5.10 cm; P=.002). Similarly, triglycerides showed the greatest reduction in the group with ≥25 interactions (mean −108.56, SD 77.06 mg/dL) compared with those with ≤15 interactions (mean −44.49, SD 50.85 mg/dL; P=.02). This group also exhibited the largest reduction in diastolic BP (mean −10.33, SD 7.40 mm Hg) compared with those with ≤15 interactions (mean −0.83, SD 7.83 mm Hg; P=.004), and the most substantial decrease in fasting glucose levels (mean −18.60, SD 10.82 mg/dL) compared with those with ≤15 interactions (mean −2.49, SD 27.54 mg/dL; P=.02). Participants in the highest quartile of digital engagement had a 60% greater likelihood of MetS reversal compared with those in the lowest quartile.

Conclusions: This study shows that combining GLP-1 and dual GIP/GLP-1 agonists with a digital behavioral change model significantly improves MetS markers in individuals with obesity. Tirzepatide proved more effective than semaglutide, leading to greater reductions in WC and triglyceride levels, along with better improvements in fasting glucose, BP, and lipid profiles. Higher app-based engagement was linked to better health outcomes, with participants in the highest engagement group having a 60% greater likelihood of treating MetS compared with those with the lowest engagement.

(Interact J Med Res 2025;14:e63079) doi:10.2196/63079

KEYWORDS

metabolic syndrome; obesity; GLP-1 medications; hybrid model of care; digital health; effectiveness; digital engagement; hybrid care; adult; cardiovascular disease; type 2 diabetes; insulin resistance; efficacy; behavioral change; obese; zone health; weight loss; monitoring; tirzepatide; semaglutide; treatment; medication; telehealth; health informatics; glucagon-like peptide-1

Introduction

Metabolic syndrome (MetS) represents a complex and multifaceted health condition characterized by a clustering of interconnected metabolic abnormalities, including central obesity, insulin resistance, dyslipidemia, and hypertension [1]. Individuals diagnosed with MetS face an elevated risk of developing cardiovascular disease [2], type 2 diabetes [3], and other related complications [4]. Understanding the intricate mechanisms underlying MetS is crucial for optimizing patient management and outcomes. In the United Arab Emirates, the prevalence of MetS was 33.6% (269/801) in the Emirati population, 34.5% (214/620) in the Arab non-Emirati population, and 40.7% (695/1709) in the Asian non-Arab population [5].

Obesity serves as a primary contributing factor to the development and progression of MetS. Excess adipose tissue, particularly visceral fat, contributes to chronic low-grade inflammation and the release of adipokines, cytokines, and free fatty acids, all of which play vital roles in insulin resistance and metabolic dysregulation [6]. Central obesity, characterized by an accumulation of adipose tissue around the abdomen, is particularly indicative of MetS and poses a heightened risk for cardiovascular complications [7]. The management of MetS revolves around lifestyle modifications and pharmacological interventions aimed at addressing its components and reducing overall cardiovascular risk [8]. Lifestyle interventions, including dietary changes, regular physical activity, and weight management, form the cornerstone of treatment [9]. In patients with MetS, dietary strategies emphasizing a balanced intake of macronutrients and lifestyle modification such as smoking cessation, regular exercise, and proper eating habits may improve profiles of each component of MetS and reduce the risk of developing diabetes and cardiovascular disease [9].

Pharmacotherapy plays a complementary role in the management of MetS, with medications targeting specific components of the syndrome. This may include statins to address dyslipidemia, antihypertensive agents to manage elevated blood pressure (BP), and insulin-sensitizing drugs to improve glucose metabolism [8]. Notably, emerging therapeutic agents such as glucagon-like peptide-1 (GLP-1) receptor agonists have shown promise in addressing multiple aspects of MetS, including glycemic control, weight reduction, and cardiovascular risk reduction [10]. By mimicking the action of endogenous GLP-1, these medications stimulate glucose-dependent insulin secretion, suppress glucagon release, and delay gastric emptying, resulting in improved glycemic control and reduced appetite [10]. Furthermore, GLP-1 receptor agonists have been associated with favorable effects on body weight, BP, and lipid profiles, making them attractive therapeutic options for individuals with MetS [10].

MetS poses a significant challenge to public health, necessitating comprehensive and more engaging approaches for prevention and management. Integrating dietary modifications, lifestyle interventions, and pharmacological treatments targeting specific components of the syndrome are essential for mitigating its adverse outcomes. The Zone.Health's meta[bolic] program, which combines pharmacotherapy with continuous engagement and monitoring to enable sustainable lifestyle modifications, demonstrated significant improvements in weight, body composition, and metabolic markers [11]. Enhancing traditional treatment paradigms by integrating continuous digital health monitoring with tailored pharmacotherapy. This program leverages advanced analytics and real time data to dynamically adjust treatment plans, fostering deeper patient engagement and more precise management of MetS components. Compared with traditional models that often rely on intermittent follow-ups and generalized treatment approaches, Zone.Health's continuous care model ensures long-term lifestyle modifications and pharmacological adherence, which are critical for long-term management of MetS. This integration of technology and personalized care is designed to significantly improve clinical outcomes by providing consistent, supportive, and adaptive interventions tailored to individual patient needs. Therefore, the aim of this study is to measure the effectiveness of GLP-1 and GLP-1/gastric inhibitory polypeptide (GIP) medications when combined with a continuous, digitally delivered behavioral change model involving a multidisciplinary team in the treatment of MetS components and weight among individuals with obesity in the UAE population.

Methods

Zone.Health Program

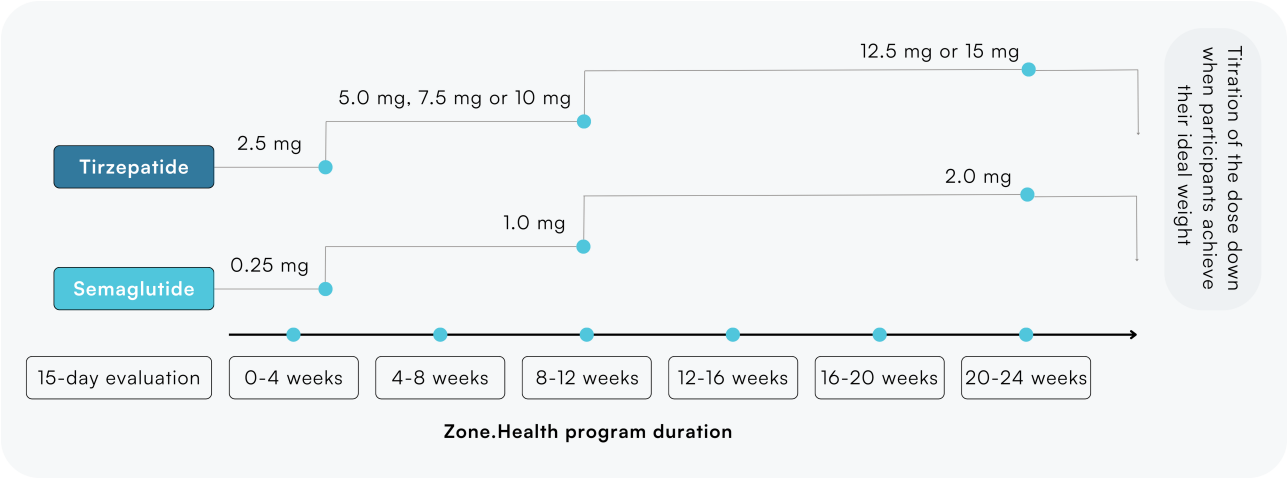
Zone.Health, introduced by meta[bolic] [12] in 2023, is a value-based 6-month weight loss program designed to be used in combination with pharmacotherapy. This innovative initiative offers hyperpersonalized insights, and monitoring in order to have patients adopt realistic, bite-sized behavioral change decisions over the duration of the program. The program's "value-based" approach underscores its commitment to efficacy, promising partial refunds to participants who fulfill specified compliance metrics but do not achieve a minimum 10% weight loss within the program's timeframe. This ensures accountability while prioritizing the participant's progress and well-being. Hyperpersonalization lies at the core of Zone.Health, where each participant's unique physiological, behavioral, and psychological profile is considered. When focusing on physiological factors, parameters like age, fat mass, muscle mass, and physical activity levels are considered. Whereas the behavioral and psychological aspect includes mental health, goal setting, and psychological beliefs. Furthermore, using continuous glucose monitoring (CGM), alongside activity and sleep tracking, and other digital biomarkers, the health care team dynamically tailors dietary recommendations, exercise plans, and medication regimens to suit each participant needs. The ongoing engagement between participants and the health care team is necessary. Biweekly sessions by the health coach comprehensive provides support, addressing progress, challenges, and evolving needs. As participants advance, engagement frequency adjusts to monthly sessions, focusing on maintaining achievements and adapting strategies for long-term success. Zone.Health's multifaceted approach extends beyond face-to-face interactions, incorporating an intuitive app called "Zone.Health" supported by the health care team including coaches, sports scientists, dietitians, physicians, and

diabetes educators. The app is integrated with a nutrition artificial intelligence food logging tool used by participants for food logging as well as integration with wearable data from Apple healthkit or Google fit. Participants are educated by dietitians to ensure correct food logging and minimize errors, by diabetes educators to interpret CGM data, and by health coaches to ensure integrating all other digital wearables (including Apple watch, Google Fitbit, or ŌURA ring [Oura Health Ltd]) into the app and educate the participants data collected from these wearables. Data integration is seamless, with unified and time-synchronized information accessible through a single clinician portal, facilitating effective patient monitoring and engagement. Medication adjustments are made monthly, guided by assessments of fat versus muscle loss ratio, weight fluctuations, and reported side effects. Regular measure of body composition analysis using (Seca mBCA 514, Seca GmbH & Co. KG) ensures a comprehensive understanding of each participant's progress, with evaluations conducted both at the program's onset and quarterly thereafter for comparison. The effectiveness of the Zone.Health's hybrid approach is supported by previous research, particularly in diabetes management, demonstrating its potential to revolutionize weight loss strategies and improve overall health outcomes [11,13,14].

Study Design

A retrospective, real world evidence observational study was conducted among participants with MetS that signed up to the Zone.Health weight loss program. A comprehensive examination of medical records was undertaken to explore the occurrence and factors contributing to MetS among patients who were prescribed GLP-1 medications. This study encompassed 51 adult patients from diverse ethnic backgrounds who had completed a 6-month program duration with all MetS parameters collected. Medication administration followed a thorough assessment, including a 15-day evaluation period and initial dietary analysis, to confirm participant eligibility. Treatment selection, whether semaglutide or tirezepatide, was determined by physicians based on clinical suitability. Following the standard practice, tirzepatide was started with an initial dose of 2.5 mg, gradually increasing to either 5.0 mg, 7.5 mg, or 10 mg by the 3-month mark and reaching 12.5 mg or 15 mg by 6 months. Semaglutide started at 0.25 mg, escalating to 1 mg at 3 months and reaching to 2.0 mg at 6 months (Figure 1). All participants adhered to the criteria, ensuring the continued validity of contractual agreements. The study strictly followed the ethical guidelines of the Declaration of Helsinki, ensuring each participant gave signed consent. The UAE health care authority monitored the clinical protocols, ensuring they met international ethical standards for research involving human subjects. All study procedures, including data collection and analysis, respected participant rights and privacy.

Figure 1. Tirzepatide and semaglutide dose titration throughout the Zone.Health program.



Study Outcomes

Primary Objective

Reduction in MetS markers (waist circumference [WC], triglycerides, and fasting glucose levels) among individuals with obesity over a 6-month period when treated with GLP-1 (semaglutide) and dual GIP/GLP-1 agonists (tirzepatide).

Secondary Objectives

First, changes in BP (systolic and diastolic) and lipid profiles (total cholesterol, low-density lipoprotein [LDL], and high-density lipoprotein [HDL]) in participants receiving tirzepatide versus those receiving semaglutide.

Second, association between levels of digital engagement with the health platform and improvements in MetS markers.

Criteria for Diagnosing MetS

International Diabetes Federation (IDF) definition of MetS was used accordingly to classify participants with MetS [15]. In this study, females and males were included in the study if they had the MetS if the following criteria were met.

First, waist circumference of ≥94 cm for males and ≥80 cm for females, based on ethnic-specific values, such as Europids, Sub-Saharan Africans, Ethnic South and Central Americans, and Middle Eastern individuals.

Second, any two of the following four factors: (1) raised triglycerides ≥150 mg/dL (1.7 mmol/L) or specific treatment

for this lipid abnormality; (2) reduced HDL cholesterol (males <40 mg/dL [1.03 mmol/L], females <50 mg/dL [1.29 mmol/L]), or specific treatment for this lipid abnormality; raised blood pressure (systolic BP ≥130 mm Hg or diastolic BP ≥85 mm Hg), or treatment for previously diagnosed hypertension; (4) raised fasting plasma glucose ≥100 mg/dL (5.6 mmol/L), or previously diagnosed type 2 diabetes.

Data Collection and Participants

Data were collected from physicians' patient records at baseline and at 6 months using the electronic medical record (Diamond, Hicom). Patients' gender, age, ethnicity, medication, WC (cm), weight kg, BP (mm Hg), and BMI (kg/m²) variables were collected. Laboratory variables were collected including total cholesterol (mg/dL), LDL cholesterol (mg/dL), HDL cholesterol (mg/dL), triglycerides (mg/dL), and fasting glucose (mg/dL). Engagement interactions were collected from the app portal. Inbound interactions refer to interactions or messages received from the participants related to dietary advice or modifications, medication dosage and side effects, coaching-related queries, and inquiries directed at any of the health care team. The outbound interactions refer to the interactions or messages initiated by the health care team to the participants, which consists of (1) biweekly reminders of food logging, weight recording, data integration, and monthly workout schedules (dietitians, coaches, and sport scientists); (2) biweekly physicians follow-up on the side effect of the medications; and (3) CGM data feedback provided by the diabetes educator.

Participant Inclusion Criteria

Eligible participants were adults (aged 18 years and older) diagnosed with MetS and obesity, who met the IDF criteria for MetS, which includes central obesity (WC specific to population and gender) plus at least 2 of the following: raised triglycerides, reduced HDL cholesterol, raised BP, or raised fasting plasma glucose. Participants had been prescribed GLP-1 or GIP/GLP-1 therapies as part of their treatment regimen. Individuals with incomplete records or those who did not engage with the digital platform were excluded from the analysis.

Ethical Considerations

Ethical approval was obtained from the Dubai Health Authority (DHA; DSREC-02/2025_09), and patient data were anonymized to maintain confidentiality. To protect participant privacy, all data collected were anonymized, and measures were taken to safeguard sensitive information, including data encryption and restricted access. It is important to note that no compensation was provided to participants for their involvement in this study.

Statistical Analysis

Data analysis performed using SPSS software version 29.0 (IBM Corp). Continuous data (such as age, weight, and laboratory parameters) were presented as means and SDs, and categorical variables (such as gender, medication type, Total n° of IDF criteria met, and ethnicity) were expressed as counts and percentage. The paired *t* test was used to compare variables at baseline and at 6 months and ANOVA test was used to see the baseline characteristics between female and male. One-way ANOVA test was conducted to test differences in baseline characteristics across medication types. ANOVA was also used to assess differences between app-based interaction groups and their impact on improvements in MetS markers. $P<.05$ was considered significant. Furthermore, post hoc Bonferroni test was conducted to measure the difference between app-based engagement groups.

Results

Basic Demographics and Baseline Characteristics Stratified by Sex (N=51)

The study population is almost evenly divided between individuals of Middle Eastern (23/51, 45%) and European (28/51, 55%) ethnicities (Table 1 and Table 2). Majority of participants are treated with tirzepatide (31/51, 61%) as opposed to semaglutide (20/51, 39%). In terms of the number of MetS parameters met, all patients met the IDF criteria with most participants (34/51, 67%) meeting 3 criteria, 27% (14/51) meeting 4, and a small group of 6% (3/51) meeting all 5 criteria. Differences between the tirzepatide and semaglutide groups are evident in several baseline characteristics. There is a significant difference in age, with the tirzepatide group being older ($P=.02$). Notable disparities also exist in weight and WC, with the tirzepatide group being significantly heavier ($P=.01$) and having a larger WC ($P=.002$). The tirzepatide group also has a higher BMI ($P=.02$). BP measurements between the groups are similar, without significant differences in either systolic or diastolic BP. Cholesterol profiles are comparable, with no significant differences in total cholesterol, LDL cholesterol, or HDL cholesterol levels. In addition, there are no significant differences in triglycerides or fasting glucose levels between the 2 groups.

Table . Basic demographics and baseline characteristics in all participants classified as having the metabolic syndrome according to the IDF[a] criteria.

Characteristic		Values (N=51), n (%)
Ethnicity		
	Middle Eastern	23 (45)
	European	28 (55)
Sex		
	Female	28 (55)
	Male	23 (45)
Total number of IDF criteria met		
	3	34 (67)
	4	14 (27)
	5	3 (6)

[a]IDF: International Diabetes Federation.

Table . Clinical and metabolic characteristics of participants on tirzepatide and semaglutide.

Characteristics	Tirzepatide (n=31), mean (SD)	Semaglutide (n=20), mean (SD)	*P* value
Age (years)	47.03 (10.49)	40.05 (9.61)	.02[a]
Weight (kg)	105.90 (26.60)	88.61 (14.23)	.01[a]
BMI (kg/m^2)	35.43 (6.90)	31.49 (3.04)	.02[a]
Systolic BP[b] (mm Hg)	127.52 (14.66)	123.55 (13.00)	.33
Diastolic BP (mm Hg)	79.03 (8.90)	76.80 (7.82)	.36
Total cholesterol (mg/dL)	198.11 (46.69)	204.52 (40.03)	.62
LDL[c] cholesterol (mg/dL)	138.39 (43.91)	149.25 (45.42)	.40
Waist circumference (cm)	114.26 (17.64)	100.33 (9.72)	.002[a]
HDL[d] cholesterol (mg/dL)	47.51 (10.61)	50.93 (11.40)	.28
Triglycerides (mg/dL)	156.41 (70.21)	171.41 (83.93)	.50
Fasting glucose (mg/dL)	106.02 (10.16)	101.72 (22.43)	.38

[a]*P* values <.05 were considered significant from 1-way ANOVA test.

[b]BP: blood pressure.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

Improvements of Diagnostic Criteria of MetS After 6 Months of Zone.Health Program

Table 3 illustrates significant improvements in the criteria used to diagnose MetS after a 6-month program, comparing the effects of tirzepatide and semaglutide (N=51). Tirzepatide showed greater reductions in WC (–18.08 cm) compared with semaglutide (−13.04 cm), both statistically significant (*P*<.001). Triglyceride levels decreased significantly with tirzepatide (−64.42 mg/dL) and even more with semaglutide (−70.70 mg/dL; *P*<.001). HDL cholesterol increases were modest and not statistically significant for both drugs. Significant reductions were observed in fasting glucose, systolic and diastolic BP, and LDL cholesterol, with tirzepatide generally showing more pronounced improvements than semaglutide. Overall cholesterol levels also decreased significantly in both treatment groups.

Table . Improvements of diagnostic criteria of metabolic syndrome after 6 months of the program in both female and male (N=51).

Variables	Tirzepatide (n=31)			Semaglutide (n=20)		
	Mean (SD)	*t* test (df)	*P* value	Mean (SD)	*t* test (df)	*P* value
Weight (kg)	−14.07 (5.83)	−13.44 (30)	<.001[a]	−13.38 (4.69)	−12.75 (19)	<.001[a]
BMI (kg/m^2)	−4.96 (2.54)	−10.69 (29)	<.001[a]	−4.53 (1.89)	−10.69 (19)	<.001[a]
Waist circumference (cm)	−18.08 (8.30)	−12.13 (30)	<.001[a]	−13.04 (7.27)	−8.02 (19)	<.001[a]
Triglycerides (mg/dL)	−64.42 (53.10)	−6.44 (29)	<.001[a]	−70.70 (69.10)	−4.58 (19)	<.001[a]
HDL[b] cholesterol (mg/dL)	2.50 (8.47)	1.68 (30)	.10	1.97 (8.67)	1.01 (19)	.32
Fasting glucose (mg/dL)	−15.32 (11.46)	−7.07 (27)	<.001[a]	−10.25 (23.64)	−1.89 (18)	.07
Systolic BP[c] (mm Hg)	−14.74 (13.92)	−5.90 (30)	<.001[a]	−14.75 (10.76)	−6.12 (19)	<.001[a]
Diastolic BP (mm Hg)	−7.23 (9.20)	−4.37 (30)	<.001[a]	−6.00 (8.32)	−3.22 (19)	.004
LDL[d] cholesterol (mg/dL)	−24.46 (32.58)	−4.04 (28)	<.001[a]	−15.15 (23.64)	−1.86 (18)	.08
Cholesterol (mg/dL)	−28.87 (45.02)	−3.51 (29)	.001[a]	−20.09 (39.98)	−2.13 (17)	.05[a]

[a]*P* values <.05 were considered significant from paired sample *t* test.

[b]HDL: high density lipoprotein.

[c]BP: blood pressure.

[d]LDL: low-density lipoprotein.

Treatment of Remaining MetS Markers in Patients With Increased WC (n=26)

In participants who did not reverse MetS due to persistently increased WC, improvements were observed with the remaining MetS markers. For instance, 85% (22/26) of participants successfully treated their elevated triglyceride levels to within normal ranges. HDL saw a significant improvement with 73% (19/26) of participants elevating their levels to the desired range. In addition, BP was stabilized in 89% (23/26) of these participants, and blood sugar levels were brought back to healthy levels in 96% (25/26) of those who did not treat MetS due to persistent increased WC.

Change in MetS Risk Factors by App-Based Interaction Groups

Table 4 and Table 5 presents the difference between patient engagement and improvements in MetS markers over a 6-month period among all participants. Engagement is quantified as both inbound and outbound interactions; inbound interactions (received from participants) average at 28.86 (SD 21.32), outbound (sent to participants from the health care team) at 58.57 (SD 25.26), and the total number of interactions at 84.00 (SD 45.76). Significant improvements were observed in several markers. WC, triglycerides, diastolic BP, and fasting glucose showed significant changes, with *P* values of .003, .01, .003, and .03, respectively. Changes in HDL and systolic BP were profound but not statistically significant. The post hoc analysis showed significant differences among interaction groups, indicating that higher app-based interactions are associated with better outcomes (Table 6). A significantly increased reduction in WC was found in individuals with ≥25 interactions (mean −19.04, SD 7.40) compared with individuals with ≤15 interactions (mean −9.60, SD 5.10 cm, *P*=.002). In addition, individuals with ≥25 interactions showed a significant reduction in triglycerides (mean −108.56, SD 77.06 mg/dL) compared with those with ≤15 interactions (mean −44.49, SD 50.85 mg/dL, *P*=.02) and a significant reduction when compared with those with 16 - 24 interactions (mean −55.77, SD 44.52 mg/dL, *P*=.03). Diastolic BP also significantly decreased in individuals with ≥25 interactions (mean −10.33, SD 7.40 mm Hg) compared with those with ≤15 interactions (mean −0.83, SD 7.83 mm Hg, *P*=.004). Fasting glucose levels were significantly lower in individuals with ≥25 interactions (mean −18.60, SD 10.82 mg/dL) compared with those with ≤15 interactions (mean −2.49, SD 27.54 mg/dL, *P*=.02).

Table . Distribution of engagement interactions by category.

Engagement interactions and category	Mean (SD)
Inbound	28.86 (21.32)
Outbound	58.57 (25.26)
Total	84.00 (45.76)

Table . Change in metabolic syndrome risk factors by app-based interaction groups (N=51).

Improvements in MetS[a] markers	Inbound app-based interactions			
	≤15 (n=12)	16-24 (n=12)	≥25 (n=27)	*P* value
WC[b] (cm)	−9.60 (5.10)	−16.01 (9.20)	−19.04 (7.40)	.003[c]
Triglycerides (mg/dL)	−44.49 (50.85)	−55.77 (44.52)	−108.56 (77.06)	.01[c]
HDL[d] (mg/dL)	+1.69 (10.03)	+2.93 (7.33)	+3.15 (5.63)	.85
Systolic BP[e] (mm Hg)	−12.58 (10.39)	−14.33 (12.79)	−15.89 (13.78)	.76
Diastolic BP (mm Hg)	−0.83 (7.83)	−4.58 (9.29)	−10.33 (7.40)	.003[c]
FG[f] (mg/dL)	−2.49 (27.54)	−13.41 (8.88)	−18.60 (10.82)	.03[c]

[a]MetS: metabolic syndrome.

[b]WC: waist circumference.

[c]*P* values <.05 were considered significant from 1-way ANOVA test.

[d]HDL: high-density lipoprotein.

[e]BP: blood pressure.

[f]FG: fasting glucose.

Table . Post hoc analysis of metabolic syndrome markers by interaction group (N=51).

Metabolic syndrome markers and interaction group (I)	Interaction group (J)	Mean difference (I–J)	*P* value
WC[a] (cm)			
≤15	16 - 24	6.41	.12
≤15	≥25	9.44[b]	.002[b]
16 - 24	≤15	−6.42	.12
16 - 24	≥25	3.03	.73
≥25	≤15	−9.44[b]	.002[b]
≥25	16 - 24	−3.03	.73
Triglycerides (mg/dL)			
≤15	16 - 24	11.28	≥.99
≤15	≥25	64.07[b]	.02[b]
16 - 24	≤15	−11.28	≥.99
16 - 24	≥25	52.79[b]	.03[b]
≥25	≤15	64.07[b]	.02[b]
≥25	16 - 24	−52.79[b]	.03[b]
Diastolic BP[c] (mm Hg)			
≤15	16 - 24	3.75	.76
≤15	≥25	9.50[b]	.004[b]
16 - 24	≤15	−3.75	.76
16 - 24	≥25	5.75	.13
≥25	≤15	−9.50[b]	.004[b]
≥25	16 - 24	−5.75	.13
FG[d] (mg/dL)			
≤15	16 - 24	10.92	.35
≤15	≥25	16.10[b]	.02[b]
16 - 24	≤15	−10.92	.35
16 - 24	≥25	5.19	≥.99
≥25	≤15	−5.19	≥.99
≥25	16 - 24	−16.10[b]	.02[b]

[a]WC: waist circumference.

[b]Mean difference is considered significant at .05 from Bonferroni post hoc test.

[c]BP: blood pressure.

[d]FG: fasting glucose.

Discussion

Principal Findings

This study examines the effect of GLP-1 and dual GIP/GLP-1 agonists when combined and integrated with a tech-enhanced, continuous feedback cycle on lifestyle modifications in treating MetS. MetS is intricately linked to central obesity and excessive weight, with studies illustrating that increased adiposity, particularly in the abdominal region, is a major risk factor for developing MetS [16]. Central obesity is associated with increased insulin resistance, which plays a pivotal role in the pathogenesis of MetS [16].

Studies have shown that even a modest amount of weight loss is associated with improvement in MetS parameters and related cardiovascular disease risk components. Losing weight is associated with a decrease in insulin resistance and insulin level, decreasing the risk of type 2 diabetes or improving previously diagnosed diabetes [17]. Losing weight also causes a positive impact on dyslipidemia by an increase in the HDL level and a drop in the LDL level [18,19]. GLP-1 and GIP agonist use has

been shown to reduce MetS severity, abdominal obesity, and inflammation [20].

Over a 6-month period, the comparative analysis of tirzepatide and semaglutide in the Zone.Health program showed significant improvements in MetS markers. Tirzepatide led to greater reductions in WC and LDL cholesterol, with slightly better results in fasting glucose and BP than semaglutide. While both treatments modestly increased HDL cholesterol without significance, semaglutide had a slightly better effect on lowering triglycerides. This suggests the treatment choice may vary based on individual patient needs and goals.

Despite significant improvements in WC observed in the study, it is noted that the rate of change might be perceived as low relative to other MetS markers. This could be attributed to several factors. First, central obesity, characterized by high WC, is often more stubborn and resistant to reduction through conventional weight loss measures alone [21]. This resistance could be partly due to the physiological characteristics of adipose tissue in the abdominal area, which is more metabolically active and sensitive to hormonal influences [21]. This may be attributed to the physiological characteristics of abdominal adipose tissue, which is highly metabolically active and hormonally sensitive, making it particularly challenging to reduce through pharmacotherapy and lifestyle changes alone. In addition, the IDF criteria consider an increased WC as a mandatory component for the diagnosis of MetS, which could preserve the perceived low reversal rate in study participants despite other significant improvements. The other MetS markers such as triglycerides, fasting glucose, and BP showed more substantial improvements even with persistent increased WC. These markers often respond more dynamically to metabolic improvements induced by GLP-1 agonists and lifestyle interventions. GLP-1 agonists improve insulin sensitivity and secretion, which directly impacts glucose control and lipid metabolism [18,19]. This improvement in insulin dynamics could explain the significant decrease in triglycerides and fasting glucose [22]. Similarly, the improvements in BP may be due to the weight loss itself as well as the improvement in arterial stiffness and endothelial function resulting from better metabolic control [23]. Furthermore, the baseline WC of our participants was significantly higher than the general thresholds for MetS, suggesting that longer-term or more intensive treatment strategies might be necessary to achieve meaningful reductions in WC. Challenges in reducing WC in this study align with findings from the Look AHEAD trial, where intensive lifestyle interventions resulted in significant weight loss but less pronounced reductions in WC [24]. This discrepancy highlights the complexity of central obesity as a metabolic entity and underscores the need for targeted strategies that specifically address this component of MetS.

Integrating Digital Solutions in Weight Loss Management and MetS Markers

The integration of hybrid care models in chronic disease management, such as diabetes mellitus, has shown promising results in reducing key metrics like HbA1c levels [11,13,14]. Our study also explores the perspective of people with MetS receiving treatment with medication such as semaglutide or tirzepatide along with feedback relating to their health in a real world environment. The long-term treatment of obesity, diabetes, and other chronic diseases requires that health care professionals shift away from episodic care and keep patients continuously engaged. The integration of digital tools as part of a standard care pathway allows health care practices to monitor and communicate with patients by observing personal data received from patients, whether it be activity, sleep habits, or information on dietary choices. Together, they are collectively able to take action regarding medication titration or a more consolidated plan during treatment [25]. The association between patient engagement and improvements in MetS markers, as presented in Table 3 of our study, further underscores the potential of digital health solutions in chronic disease management, particularly when combined with pharmacotherapy. Our findings indicate that higher engagement in inbound interactions is linked to better outcomes in key MetS parameters such as WC, triglycerides, and diastolic BP, and are supported by previous studies around engagement and outcomes in diabetes [13]. The significant reduction in these MetS markers with increased digital interaction suggests that patient engagement may be as critical as medication adherence in the management of MetS. This aligns with the broader trend in health care that emphasizes patient-centered models where ongoing patient engagement and personalized interaction play pivotal roles. The integration of digital tools facilitates this by providing continuous monitoring and real time feedback, which are essential for sustaining behavior changes over time. Furthermore, the Peterson Report emphasizes that purely digital solutions often fall short in achieving long-term patient engagement and effective disease management, citing a lack of personalization and direct human interaction as key shortcomings [26]. This finding supports the adoption of a hybrid care model, which combines the scalability of digital tools with the tailored support and expertise of the treating health care professionals.

To our knowledge, this study is the first to examine the effectiveness of GLP-1 medications alongside lifestyle interventions under a hybrid model of care in improving MetS markers. However, this study has a few limitations. First, the sample size, although adequate for detecting significant changes, is not large enough to ensure the generalizability of the findings across diverse populations or to detect more subtle effects. Second, the absence of a control group makes it difficult to definitively attribute the improvements in MetS markers to the interventions, as comparisons with a nonintervention baseline are lacking. In addition, Zone.Health is a paid program, and monetary investments into the program may have influenced participant outcomes, as those who have paid for the program might be more likely to participate actively and follow instructions. What can be concluded is that GLP outcomes on certain MetS parameters can be enhanced when delivered in a hybrid care model, as demonstrated in previous studies, thereby leading to better outcomes, an important factor considering the cost of GLP treatments.

Conclusions

In conclusion, this study highlights the effective role of combining GLP-1 and GIP/GLP-1 agonists with a continuous,

digitally delivered behavioral change model by an integrated care team in managing MetS, particularly in reducing WC, improving lipid profiles, and enhancing glycemic control across different genders. Despite notable improvements in most MetS markers, the persistence of increased WC suggests a need for strategies specifically targeted at abdominal obesity. Future research should focus on optimizing interventions that specifically address this resistant aspect of MetS, potentially incorporating new digital biomarker monitoring to enhance engagement even further due to the relationship between engagement and outcomes, as shown in this study.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HZ, HJ, JK, and IA contributed to the conceptualization. HZ, SA, NA, and IA contributed to the methodology. HZ, SA, HJ, JD, and IA contributed to the investigation. Project administration was handled by HZ. HZ and IA contributed to the visualization. HZ and SA contributed to data curation. HZ, HJ, SA, and IA contributed to the writing – original draft. MC, JD, YS, JK, NA, and AH contributed to the writing – review and editing.

Conflicts of Interest

The following authors are full-time employees or interns at GluCare: HZ, HJ, SA, MC, JD, YS, JK, and NA. The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript: AH and IA.

References

1. Han TS, Lean MEJ. Metabolic syndrome. Medicine (Abingdon) 2015 Feb;43(2):80-87. [doi: 10.1016/j.mpmed.2014.11.006]
2. Guembe MJ, Fernandez-Lazaro CI, Sayon-Orea C, Toledo E, Moreno-Iribas C, RIVANA Study Investigators. Risk for cardiovascular disease associated with metabolic syndrome and its components: a 13-year prospective study in the RIVANA cohort. Cardiovasc Diabetol 2020 Nov 22;19(1):195. [doi: 10.1186/s12933-020-01166-6] [Medline: 33222691]
3. Shin JA, Lee JH, Lim SY, et al. Metabolic syndrome as a predictor of type 2 diabetes, and its clinical interpretations and usefulness. J Diabetes Investig 2013 Jul 8;4(4):334-343. [doi: 10.1111/jdi.12075] [Medline: 24843675]
4. Alberti K, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 2009 Oct 20;120(16):1640-1645. [doi: 10.1161/CIRCULATIONAHA.109.192644] [Medline: 19805654]
5. Mahmoud I, Sulaiman N. Prevalence of metabolic syndrome and associated risk factors in the United Arab Emirates: a cross-sectional population-based study. Front Public Health 2021;9:811006. [doi: 10.3389/fpubh.2021.811006] [Medline: 35141192]
6. Kawai T, Autieri MV, Scalia R. Adipose tissue inflammation and metabolic dysfunction in obesity. Am J Physiol Cell Physiol 2021 Mar 1;320(3):C375-C391. [doi: 10.1152/ajpcell.00379.2020] [Medline: 33356944]
7. Chait A, den Hartigh LJ. Adipose tissue distribution, inflammation and its metabolic consequences, including diabetes and cardiovascular disease. Front Cardiovasc Med 2020;7:22. [doi: 10.3389/fcvm.2020.00022] [Medline: 32158768]
8. Rask Larsen J, Dima L, Correll CU, Manu P. The pharmacological management of metabolic syndrome. Expert Rev Clin Pharmacol 2018 Apr;11(4):397-410. [doi: 10.1080/17512433.2018.1429910] [Medline: 29345505]
9. Kim HL, Chung J, Kim KJ, et al. Lifestyle modification in the management of metabolic syndrome: statement from Korean Society of CardioMetabolic Syndrome (KSCMS). Korean Circ J 2022 Feb;52(2):93-109. [doi: 10.4070/kcj.2021.0328] [Medline: 35128848]
10. Seon MJ, Hwang SY, Son Y, Song J, Kim OY. Circulating GLP-1 levels as a potential indicator of metabolic syndrome risk in adult women. Nutrients 2021 Mar 6;13(3):865. [doi: 10.3390/nu13030865] [Medline: 33800785]
11. Zakaria H, Alshehhi S, Caccelli M, et al. Effectiveness of a hybrid approach in integrating GLP-1 agonists and lifestyle guidance for obesity and pre-diabetes management: RWE retrospective study. Metabol Open 2024 Jun;22:100283. [doi: 10.1016/j.metop.2024.100283] [Medline: 38699398]
12. Meta[bolic]. URL: https://metabolic.health/ [accessed 2025-03-24]
13. Zakaria H, Said Y, Aleabova S, et al. Measuring the effectiveness of hybrid diabetes care over 90 days through continuous data monitoring in type 2 diabetic patients. Front Endocrinol (Lausanne) 2024;15:1355792. [doi: 10.3389/fendo.2024.1355792] [Medline: 38774233]
14. Almarzooqi I, Zakaria H, Aleabova S, et al. Effectiveness of a hybrid technology enabled care model as measured by ICHOM standard set on established and managed type 2 diabetes already using medications: a RWE retrospective study. Metabol Open 2023 Dec;20:100262. [doi: 10.1016/j.metop.2023.100262] [Medline: 38115869]

15. Zimmet P, M M Alberti KG, Serrano Ríos M. A new international diabetes federation worldwide definition of the metabolic syndrome: the rationale and the results. Rev Esp Cardiol 2005 Dec;58(12):1371-1376. [Medline: 16371194]
16. Grundy SM. Metabolic syndrome pandemic. Arterioscler Thromb Vasc Biol 2008 Apr;28(4):629-636. [doi: 10.1161/ATVBAHA.107.151092] [Medline: 18174459]
17. Yamakage H, Jo T, Tanaka M, et al. Five percent weight loss is a significant 1-year predictor and an optimal 5-year cut-off for reducing the number of obesity-related cardiovascular disease risk components: the Japan obesity and metabolic syndrome study. Front Endocrinol (Lausanne) 2024;15:1343153. [doi: 10.3389/fendo.2024.1343153] [Medline: 38601201]
18. Williams PT, Stefanick ML, Vranizan KM, Wood PD. The effects of weight loss by exercise or by dieting on plasma high-density lipoprotein (HDL) levels in men with low, intermediate, and normal-to-high HDL at baseline. Metab Clin Exp 1994 Jul;43(7):917-924. [doi: 10.1016/0026-0495(94)90277-1] [Medline: 8028519]
19. Khan AA, Mundra PA, Straznicky NE, et al. Weight loss and exercise alter the high-density lipoprotein lipidome and improve high-density lipoprotein functionality in metabolic syndrome. Arterioscler Thromb Vasc Biol 2018 Feb;38(2):438-447. [doi: 10.1161/ATVBAHA.117.310212] [Medline: 29284607]
20. Sandsdal RM, Juhl CR, Jensen SBK, et al. Combination of exercise and GLP-1 receptor agonist treatment reduces severity of metabolic syndrome, abdominal obesity, and inflammation: a randomized controlled trial. Cardiovasc Diabetol 2023 Feb 25;22(1):41. [doi: 10.1186/s12933-023-01765-z] [Medline: 36841762]
21. Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. Nature New Biol 2006 Dec 14;444(7121):881-887. [doi: 10.1038/nature05488] [Medline: 17167477]
22. Nauck MA, Meier JJ, Cavender MA, Abd El Aziz M, Drucker DJ. Cardiovascular actions and clinical outcomes with glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. Circulation 2017 Aug 29;136(9):849-870. [doi: 10.1161/CIRCULATIONAHA.117.028136] [Medline: 28847797]
23. Landsberg L, Aronne LJ, Beilin LJ, et al. Obesity-related hypertension: pathogenesis, cardiovascular risk, and treatment--a position paper of the The Obesity Society and The American Society of Hypertension. Obesity (Silver Spring) 2013 Jan;21(1):8-24. [doi: 10.1002/oby.20181] [Medline: 23401272]
24. Dutton GR, Lewis CE. The Look AHEAD trial: implications for lifestyle intervention in type 2 diabetes mellitus. Prog Cardiovasc Dis 2015;58(1):69-75. [doi: 10.1016/j.pcad.2015.04.002] [Medline: 25936906]
25. Rhee SY, Kim C, Shin DW, Steinhubl SR. Present and future of digital health in diabetes and metabolic disease. Diabetes Metab J 2020 Dec;44(6):819-827. [doi: 10.4093/dmj.2020.0088] [Medline: 33389956]
26. Digital Diabetes Management Solutions: Peterson Health Technology Institute (PHTI); 2024.

Abbreviations

BP: blood pressure
CGM: continuous glucose monitoring
GIP: gastric inhibitory polypeptide
GLP-1: glucagon-like peptide-1
HDL: high-density lipoprotein
IDF: International Diabetes Federation
LDL: low-density lipoprotein
MetS: metabolic syndrome
WC: waist circumference

Edited by TDA Cardoso; submitted 10.06.24; peer-reviewed by AJ Graybeal, L Qian; revised version received 20.02.25; accepted 24.02.25; published 27.03.25.

Please cite as:
Zakaria H, Jabri H, Alshehhi S, Caccelli M, Debs J, Said Y, Kattan J, Almarzooqi N, Hashemi A, Almarzooqi I
Glucagon-Like Peptide-1 Receptor Agonists Combined With Personalized Digital Health Care for the Treatment of Metabolic Syndrome in Adults With Obesity: Retrospective Observational Study
Interact J Med Res 2025;14:e63079
URL: https://www.i-jmr.org/2025/1/e63079
doi:10.2196/63079

Interactive Effects of Weight Recording Frequency and the Volume of Chat Communication With Health Care Professionals on Weight Loss in mHealth Interventions for Noncommunicable Diseases: Retrospective Observational Study

Yuta Hagiwara[1], MSci; Takuji Adachi[2], PhD; Masashi Kanai[1,3], PhD; Kotoe Shimizu[1], BSc; Shinpei Ishida[1], MSci; Takahiro Miki[1], PhD

[1]PREVENT Inc, IKKO Shinsakae Building, 1 Aoi, Higashi-ku, Aichi, Japan

[2]Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Aichi, Japan

[3]Institute of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan

Corresponding Author:
Takahiro Miki, PhD
PREVENT Inc, IKKO Shinsakae Building, 1 Aoi, Higashi-ku, Aichi, Japan

Abstract

Background: Mobile health (mHealth) apps are increasingly used for health promotion, particularly for managing noncommunicable diseases (NCDs) through behavior modification. Understanding the factors associated with successful weight loss in such interventions can improve program effectiveness.

Objective: This study examined factors influencing weight change and the relationship between weight recording frequency and chat volume with health care professionals on weight loss in individuals with obesity and NCDs.

Methods: The participants had obesity (BMI ≥25 kg/m²) and were diagnosed with NCDs (eg, hypertension, diabetes, dyslipidemia). The program included 12 telephone consultations with health care professionals. Only participants who completed the full 6-month program, including all 12 telephone consultations, and provided an end-of-study weight were included in the analysis. The primary outcome was the rate of weight change, defined as the percentage change in weight from the initial period (first 14 days) to the final period (2 weeks before the last consultation), relative to the initial weight. The key independent variables were proportion of days with weight recording and chat communication volume (total messages exchanged). An interaction term between these variables was included to assess moderation effects in the regression analysis. The volume of communication was measured as the total number of messages exchanged, with each message, regardless of who sent it, being counted as 1 interaction. Health care staffs were instructed to send a single scheduled chat message per week following each biweekly phone consultation. These scheduled messages primarily included personalized feedback, reminders, and motivational support. In addition, providers responded to participant-initiated messages at any time during the program. Furthermore, 1 professional responded to each participant. Hierarchical multiple regression and simple slope analyses were conducted to identify relationships and interactions among these variables.

Results: The final analysis of this study included 2423 participants. Significant negative associations were found between the rate of weight change and baseline BMI (β=−.10; *P*<.001), proportion of days with weight recording (β=−.017; *P*<.001), and communication volume (β=−.193; *P*<.001). The interaction between proportion of days with weight recording and chat frequency also showed a significantly negative effect on weight change (β=−.01; *P*<.001). Simple slope analysis showed that when the proportion of days with weight recording was +1 SD above the mean, frequent chats were associated with greater weight reduction (slope=−0.60; *P*<.001), whereas no significant effect was observed at −1 SD (slope=−0.01; *P*=.94)

Conclusions: The findings suggest that both the proportion of days with weight recording and communication volume independently and interactively influence weight change in individuals with obesity and NCDs.

(Interact J Med Res 2025;14:e65863) doi:10.2196/65863

KEYWORDS

weight change; behavior modification; health care communication; weight recording; chat communication; text communication; health care professionals; weight loss; mHealth; mobile health; app; digital health; smartphone; mobile health intervention; noncommunicable disease; NCD; weight loss outcome; obesity; overweight; retrospective study; observational study; cerebrovascular disease; cardiovascular disease; lifestyle modification; mobile phone

Introduction

Obesity is a major public health challenge and a well-established risk factor for noncommunicable diseases (NCDs), such as hypertension, diabetes, and dyslipidemia. Effective weight management is crucial not only for obesity prevention but also for reducing the risk of NCD-related complications. Obesity is recognized as one of the contributing factors to the onset of NCDs, such as ischemic heart disease, cerebrovascular disease, hypertension, and diabetes mellitus. It was found that 43% of individuals aged 18 years and older were overweight and 16% were living with obesity in 2022 over the world [1,2]. Several guidelines for primary and secondary prevention of NCDs recommend weight loss for managing risk factors derived from insulin resistance [3-6]. Interventions causing any weight loss significantly reduce systolic and diastolic blood pressure, low-density lipoprotein cholesterol, triglycerides, fasting plasma glucose, and HbA_{1c} in a period of over 6 - 12 months, and lasting benefits persist for at least 2 years [7]. Furthermore, other studies have shown that patients with diabetes who achieved ≥10% weight loss in the first year after diagnosis had a significantly higher likelihood of remission [8]. Thus, effective strategies for weight reduction are important for managing and preventing NCDs.

The use of mobile health (mHealth) apps is an emerging approach, which has been proven to promote weight loss through lifestyle modifications [9]. One of the key functions of mobile apps is daily lifestyle records and health status monitoring, which empower individuals to actively manage their health and enhance efficiency for health care providers [10]. These apps offer tools for self-monitoring, allowing users to record their weight, dietary intake, and physical activity. Study has confirmed the effectiveness of self-monitoring in promoting weight loss [11], with frequent weight recording being particularly associated with greater weight loss [12,13].

Recently, interactive approaches that include communication with health care professionals are considered highly effective in mHealth intervention [14,15]. These involve phone consultations, chat messaging, and personalized feedback, supporting swifter information exchange between the user and health care professionals. Studies have also shown that personalized advice and social support from health care professionals enhance adherence to weight loss programs and improve outcomes [14,15]. The success of these mHealth interventions is due to user engagement, specifically, the frequency of users recording their data in the app. High engagement rates are crucial for effective self-monitoring and allow health care professionals to provide relevant feedback.

Various opinions exist on what constitutes the most effective form of interactive support within these programs. Although some argue for the quality of interactions, such as the content and personalization of messages, others have focused on the importance of the quantity of interactions. Our review of the literature suggests that the volume of communication between users and health care professionals could be a critical factor in the effectiveness of weight loss interventions [16-20]. However, only limited evidence can confirm the interactive effect of adherence to self-monitoring and the quantity of app-based communication volume on achieving weight reduction.

Therefore, this study aims to investigate the relation of weight recording frequency and app-based communication volume with weight reduction and determine their interactive effects on the participants of the app-based lifestyle modification.

Methods

Study Design and Setting

This retrospective observational study used data from PREVENT Inc., a company providing medical data analysis and mobile app-based lifestyle modification support programs for individuals with NCDs. PREVENT Inc. predicts the risk of the onset of cerebrovascular and cardiovascular diseases using health insurance claims and health checkup data. Furthermore, employees or their dependents who are at high risk of onset could be provided with an individualized support program by health care professionals.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki, and was approved by the Research Ethics Committee of Nagoya University (approval 2022 - 0453). Participants agreed to a privacy policy at the start of the program, which stated that the data gathered in the app may be used for future study. All participants provided informed consent before participating in the study. No financial compensation was provided to participants, who were free to withdraw at any time. All data presented were anonymized, and there are no images in the paper or the Multimedia Appendix that could identify individual participants.

Study Population and Procedure

PREVENT, Inc provides a mobile app-based lifestyle modification support program called Mystar to individuals at a high risk of cerebrovascular and cardiovascular diseases with NCDs, commissioned by the health insurance association. The selection of the target population was performed objectively using health insurance claims and health checkup data provided by the health insurance association. The principal approach to participation in these targeted population programs was for the health insurance association or PREVENT, Inc to send an invitation to participate in the program, or for the health insurance association to contact the individual. This study analyzed data from individuals who participated in the program from December 2018 to November 2023.

Participants eligible for this study were required to meet specific inclusion criteria. These criteria included a diagnosis of hypertension, diabetes, or dyslipidemia, with participants either receiving medication or having a history of coronary artery disease or stroke. Furthermore, participants needed to be classified as having obesity, defined by a BMI of 25 kg/m^2 or higher. Only individuals who agreed to participate in the 6-month program and provided informed consent for the use of their data in study were included. Exclusion criteria were applied to ensure the integrity of the study. Participants who failed to complete the full 6-month program (dropouts) or whose BMI

was below 25 kg/m² at the time of enrollment were excluded. In addition, individuals with medical conditions that could interfere with participation, such as severe cardiac diseases (eg, arrhythmia or cardiomyopathy), end-stage kidney disease, or mental disorders, were not eligible. The study also excluded participants who were taking medications that might affect weight management interventions, such as cardiotonic agents, immunosuppressive drugs, or antipsychotics. Finally, participants deemed unsuitable for inclusion by the health insurance association were excluded based on predefined criteria.

Disease Management Program

The disease management program was implemented with the approval of the participants' attending physicians. The program's objective was to improve the management of certain conditions, such as hypertension, diabetes, and dyslipidemia, through lifestyle improvements, to prevent the onset of conditions, such as stroke and myocardial infarction. Participants were provided with a mobile app account, which allowed them to record their lifestyle habits and exchange chat messages with health care professionals assigned to them. The program lasted 6 months, during which the participants had 12 telephone consultations with health care professionals. Nurses, registered dietitians, physical therapists, and other health care professionals provided individual advice through phone calls and chats based on the participants' lifestyle and disease management conditions. During telephone consultations, the participants set behavioral goals for the next 2 weeks and worked on improving lifestyle habits, such as exercise and diet by recording self-monitoring of health behaviors. Using the chat feature, not only could the participants ask health care professionals questions anytime but health care professionals could also provide individual advice. Further details of the program are described in a previous study [21].

Measurement of the Study Outcome: Weight Change

This study's primary outcome was that the rate of weight loss because the rate of weight change has been shown to be important for health improvement in individuals with obesity in Japan [22]. The rate of weight loss was calculated as the percentage change in weight over the program duration. Specifically, it was determined by dividing the difference between the initial average weight (recorded during the first 14 days of the program) and the final average weight (recorded during the 2 weeks before the last consultation) by the initial average weight, and then multiplying by 100 to express it as a percentage. This approach ensures that weight change is represented as a proportion of the initial weight, rather than just an absolute difference.

Proportion of Days With Weight Recording

Weight recording was measured as the proportion of days during the program on which participants recorded their weight at least once. If a participant recorded their weight more than once on a single day, only the first entry was used to calculate daily adherence. Units are reported as the percentage of days with weight recording relative to the total number of days in the program.

Communication Volume

Communication volume was assessed by counting the total number of chat messages exchanged between participants and health care staff throughout the program. Each chat message, whether initiated by the participants or health care staff, was considered as 1 interaction. In addition, 1 professional responded to each participant. The participants received advice from their assigned health care staff a week after each of the biweekly phone consultations, which was implemented over 6 months, totaling 12 sessions. Chat interactions included responses to participant inquiries, personalized feedback, and reminders for weight tracking when participants failed to log their weight. Health care professionals involved in these interactions included nurses, registered dietitians, and physical therapists, and messages were tailored based on participants' conditions and progress. Apart from these scheduled messages, the participants and health care professionals could freely engage in chat exchanges at any time during the program. In these chat communications, the health care staff provided feedback, lifestyle modification guidance, and information that encouraged continued healthy behaviors based on the participants' self-monitoring of health behaviors data. The participants could report the details of their daily activities and ask the health care staff about any questions or concerns. Each chat message, whether initiated by the participants or health care staff, was considered as 1 interaction.

Other Variables

We collected comprehensive demographic data (eg, age, gender, height, weight, and BMI), medical information, self-monitoring of health behaviors data, and app usage statistics from the participants. In addition, medical history information was obtained from health insurance claims data before the program began. Demographic and medication data were collected at program enrollment, while self-monitoring and app usage data were recorded throughout the program.

Independent Variables

The key independent variables in this study were the proportion of days with weight recording and chat communication volume. In addition, age, gender, and BMI were included as independent variables to control for demographic and physiological factors. An interaction term between the proportion of days with weight recording and chat communication volume was introduced in Model 2 to examine their combined effect on weight change.

Statistical Analysis

To handle missing data, a listwise deletion method was used, ensuring that only complete cases were analyzed. Descriptive statistics were used to summarize the basic demographic and clinical characteristics of the study population, including age, gender, BMI, proportion of days with weight recording, and volume of communication with health care professionals. To assess potential selection bias, we compared baseline characteristics between participants who completed the program and those who were excluded due to missing data. Independent *t* tests were used for continuous variables, and χ^2 tests were applied for categorical variables. A hierarchical multiple regression analysis was performed to investigate the relation

between the dependent variable—the rate of weight change—and various independent variables. Model 1 included age, gender, BMI, proportion of days with weight recording, and communication volume as the independent variables. Model 2 expanded upon Model 1 by introducing an interaction term between proportion of days with weight recording and chat count to examine their combined effect on weight change. The fit of Models 1 and 2 was compared to assess the contribution of the interaction term in explaining the variability in weight change. Upon identifying a significant interaction effect, a simple slope analysis was performed to further investigate the nature of this interaction. The proportion of days with weight recording was categorized into 3 groups based on SD, which included values more than one SD above the mean (+1 SD group), the mean itself (mean group), and one SD below the mean (−1 SD group). All statistical analyses were performed using R (version 4.1.0; R Foundation for Statistical Computing).

Results

Baseline Characteristics of the Study Participants

A total of 3999 participants were assessed for eligibility, and 90 dropped out during the program. Thus, 3909 participants (97.7%) completed the 6-month programme. Of these, 1486 were excluded from the analysis due to missing data, including those who did not provide data at the end of the study. Consequently, the final analysis was restricted to 2423 participants who met the criteria for the complete-data analysis. Table 1 shows the characteristics of the 2423 participants included in this study. Baseline characteristics were compared between completers (n=2423) and excluded participants (n=1486). Participants with complete data were older, more likely to be male, and had a higher prevalence of hypertension, while excluded participants had a higher prevalence of diabetes mellitus (P<.01 for all). Participants with complete data also had lower BMI, lower HbA_{1c}, and greater app engagement (higher chat volume and proportion of days with weight recording; P<.01). Detailed results were provided in Multimedia Appendix 1.

Table . Participants' characteristics. Values are shown as mean (SD) for ordinal variables and counts (%) for categorical variables.

Participants' characteristics	Results (n=2423)
Age (years), mean (SD)	54.9 (6.8)
Gender, n (%)	
Men	2144 (88.5)
Women	279 (11.5)
NCDs[a]	
Hypertension, n (%)	1842 (76)
Diabetes mellitus, n (%)	1199 (49.5)
Dyslipidemia, n (%)	1510 (62.3)
Previous stroke, n (%)	79 (3.3)
Previous ischemic heart disease, n (%)	151 (6.2)
Condition at the start of the program, mean (SD)	
Body weight (kg)	83.4 (12.7)
BMI (kg/m^2)	29.2 (3.7)
Systolic blood pressure (mmHg)	130.9 (12.2)
Diastolic blood pressure (mmHg)	84.5 (9.4)
HbA_{1c} (%)	6.5 (1.1)
HDL[b] cholesterol (mg/dL)	51.5 (11.9)
LDL[c] cholesterol (mg/dL)	121.8 (30.3)
App-related factor, mean (SD)	
Chat volume (chats per week)	1.9 (1.8)
Proportion of days with weight recording (%)	78.6 (26.7)

[a]NCD: noncommunicable disease.

[b]HDL: high-density lipoprotein.

[c]LDL: low-density lipoprotein.

Factors Associated With the Rate of Weight Change

Hierarchical multiple regression analyses were conducted to determine the factors associated with the rate of weight change among the participants. In Model 1, the independent variables included age, gender, baseline BMI, proportion of days with weight recording, and communication volume. The analysis identified significant negative coefficients for baseline BMI (β=−.097; *P*<.001), proportion of days with weight recording(β=−.017; *P*<.001), and communication volume (β=−.193; *P*<.001). Model 2 extended Model 1 by incorporating an interaction term between proportion of days with weight recording and chat frequency, which assesses their combined effect on weight change. The results of Model 2 showed a significant negative interaction effect (β=−.006; *P*<.001), suggesting that higher proportion of days with weight recording combined with frequent chats substantially enhance weight loss. Furthermore, Models 1 and 2 have shown that the addition of the interaction term significantly enhanced model performance (Δ*F*=13.97; *P*<.001). Table 2 has more details.

Table . Results of hierarchical linear regression analysis for weight change.

	Model 1[a] (Adjusted R^2=0.06)		Model 2[b] (Adjusted R^2=0.06)	
	β	*P* value	β	*P* value
Age	.016	.07	.02	.06
Gender (reference: Men)	.398	.02	.39	.03
BMI	−.097	<.001	−.10	<.001
Proportion of days with weight recording	−.017	<.001	−.02	<.001
Volume of chat communication	−.193	<.001	−.15	<.001
Proportion of days with weight recording × volume of chat communication (interaction)	—[c]	—	−.01	<.001

[a]Model 1 includes age, gender, BMI, proportion of days with weight recording, and volume of chat communication as independent variables.

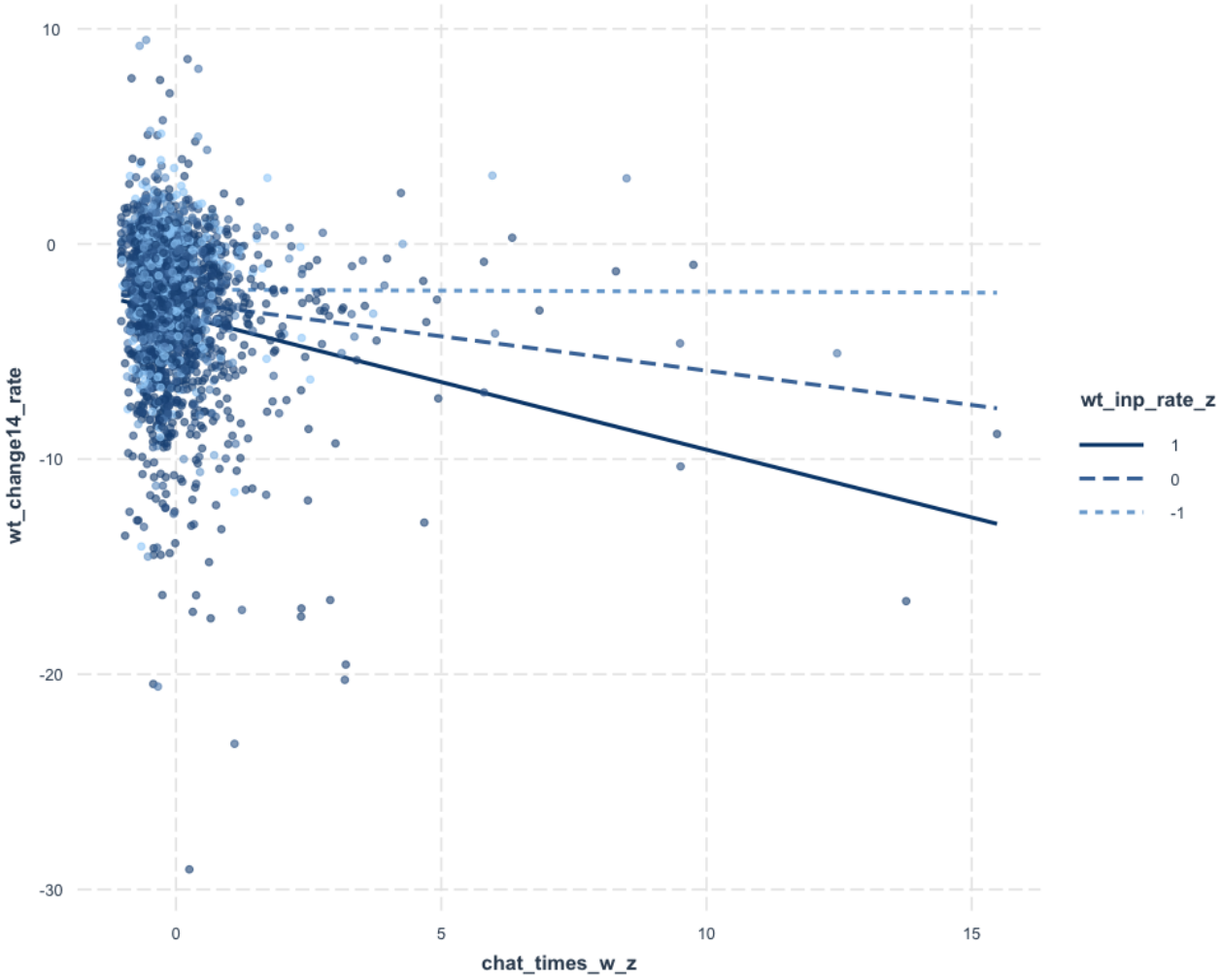
[b]Model 2 extends model 1 by adding an interaction term between proportion of days with weight recording and volume of chat communication.

[c]Not applicable.

Simple Slopes Analysis

When the proportion of days with weight recording is −1 SD below the mean, the slope of weekly chat times on weight change rate is −0.01; this result is not statistically significant (*P*=.94). When the proportion of days with weight recording is at the mean (0 SD), the slope of weekly chat times on weight change rate is −0.32; this result is significant (*P*<.001). When the proportion of days with weight recording is +1 SD above the mean, the slope of weekly chat times on weight change rate is −0.6; this result is significant (*P*<.001) (see Figure 1).

Figure 1. The figure illustrates the interaction between proportion of days with weight recording (wt_inp_rate_z) and communication volume (chat_times_w_z) on weight change rate (wt_change14_rate).



Discussion

Principal Findings

This study aims to determine the factors associated with weight change rates in mHealth and assess the interaction of proportion of days with weight recording and communication volume on weight change rates. As a result, the regression analysis indicated that chat communication volume had a stronger association with weight change rates compared with the proportion of days with weight recording. Furthermore, the higher chat communication volume was associated with greater weight reduction in the case of high proportion of days with weight recording, whereas it was not significant when the proportion of days with weight recording were low.

The primary findings of this study were that the volume of chat communication and proportion of days with weight recording were associated with the weight change rate and that a significant interactive effect on weight loss was observed during the lifestyle modification program. Chat messages included both generic reminders, such as prompts to log weight, and personalized feedback based on participant progress and health data, as reflected by the high proportion of days with weight recording and the regular chat-based communication (1.9 exchanges per week) throughout the program. These interactions, combining standardized guidance and tailored support, may have helped participants stay motivated and accountable. In this study, the average proportion of days with weight recording over 6 months was 78.6% (1902/2423), indicating a higher user engagement. The lifestyle modification program in this study required biweekly telephone consultations based on the lifestyle data including weight and provided personalized feedback. Furthermore, this interactive approach likely contributed to encouraging the participants' proportion of days with weight recording. Previous studies have reported on the relationship between each indicator, proportion of days with weight recording, and communication volume, as well as the achievement of weight loss [11,12,14,15]. Weight measurement has been considered as one of the key indicators reflecting the user's engagement, which is related to weight loss outcomes in a mobile app intervention [12]. A study suggested that regular self-weighing patterns, particularly high and consistent self-weighing, are associated with greater weight loss and maintenance, as well as better adherence to energy intake and step goals in a 12-month behavioral weight loss intervention [13]. Another study of physical activity promotion determining the effectiveness of behavior modification support through gamified mobile apps has found that users with high engagement

showed greater improvements in physical activity [23]. However, it is still unclear how the interaction between the proportion of days with weight recording and the amount of communication affects weight loss due to lifestyle modifications. In this study, the interaction between the amount of communication and the effect of weight loss was higher when the proportion of days with weight recording was high, whereas no significant relation was observed between the amount of communication and the effect of weight loss when the proportion of days with weight recording was low. These findings suggest that frequent weight recording tends to increase the impact of communication with health care professionals, making personalized feedback more effective and contributing to greater weight loss.

Mainly 2 potential reasons could explain the findings of this study. First, a high proportion of days with weight recordings likely result in a richer and more continuous stream of data, allowing health care professionals to better understand the participants' condition and provide more tailored and precise advice. Conversely, a low proportion of days with weight recording tend to provide fragmented data, possibly resulting in more general feedback that may not be as effective. A systematic review by Berry et al [24] has reported that tailored advice in digital self-monitoring of physical activities and diet is an effective intervention for promoting weight loss in adults with obesity or overweight. Another study has reported that individualized text messages and treatment information provision tend to facilitate user behavioral change and the development of healthy behaviors [25]. The assumption from this study is that if health care providers can obtain large amounts of self-monitoring of health behaviors data from participants, it will be possible to provide highly personalized health plans and advice. This concept is in line with the studies reporting that lifestyle modification was promoted more by individualized feedback, reminders, and messages compared with standardized communication [26,27]. If only limited self-monitoring of health behaviors data were available, health care professionals would be unable to provide personalized advice, and the advice given would predominantly be general. Furthermore, a second possible explanation is the theory that frequent weight monitoring will foster the habit of self-monitoring, which will motivate participants to be more consistent in achieving their weight loss goals, thereby maximizing the effectiveness of communication with their instructors. A study by Conroy et al [28] investigated the relation between physical activity self-monitoring and weight loss and found that frequent self-monitoring led to increased physical activity and weight loss. Another study has found that users who stopped tracking their weight experienced weight gain and a decline in both their physical activity and activity tracking frequency [29]. These results suggest the association between weight monitoring and the motivation to achieve lifestyle modifications. Furthermore, patient–provider communication has been reported to affect adherence through trust and motivation [30]. This heightened motivation enhances the impact of communication with health care professionals as participants are more likely to heed the advice seriously and act upon it.

The findings emphasize the critical role of self-monitoring in the context of mHealth interventions. As digital health technologies become increasingly integrated into daily life, understanding the mechanisms and effects of self-monitoring on behavioral change is important. Self-monitoring is acknowledged as a cornerstone of behavioral change including lifestyle modification [31], involves individuals in self-care management, and enhances self-feedback through learning the link between daily lifestyle changes and health outcomes. The effectiveness of weight and lifestyle monitoring for weight reduction has been widely accepted even before the development of mHealth [24]. This study on mHealth intervention has confirmed the potential role of self-monitoring in the relation between mobile app-based communication and lifestyle modification, emphasizing the unique role of self-monitoring in the mHealth era. Thus, lifestyle monitoring using mobile apps may facilitate the positive cycle of behaviors and outcomes caused by personalized communication, which is a determinant of the effectiveness of lifestyle modifications in mHealth [24,32]. However, the high proportion of days with weight recording in this study may indicate the high readiness of the users for the program or their intrinsic motivation for lifestyle modification. Furthermore, the potential burden on the participants to consistently record their weight in daily living should also be considered to monitor long-term adherence. Therefore, further study is needed to investigate the causal relation between self-monitoring passively promoted by mHealth intervention, improved health care communication with health professionals, and subsequent behavioral change. In addition, future study should explore how chat interactions influence participant retention, particularly among those who do not experience early weight loss. Understanding whether chat engagement can help maintain motivation in these individuals could provide further findings for optimising mHealth interventions for long-term adherence.

Limitations

This study performed a retrospective analysis targeting participants in health programs provided by health insurance unions. Consequently, the participants were predominantly those who voluntarily joined the programs, the majority being males. This demographic imbalance may limit the generalizability of the findings. Future study should investigate whether similar trends are observed in more diverse populations, including elderly individuals and females. In addition, our analysis was conducted using a completers-only approach, meaning that participants who dropped out before completing the program were not included in the final analysis. Since previous study has shown that individuals who drop out of weight loss interventions tend to have poorer outcomes, our findings may not fully reflect the weight loss trajectories of all participants [33]. In fact, baseline characteristics differed significantly between participants with complete data and those with missing data. These differences suggest that participants with complete data may have been more intrinsically motivated or in better health, potentially influencing weight loss outcomes. Furthermore, this study measured the communication volume solely based on the number of chat messages sent and received, without considering the length or content of the messages. While

message frequency is a useful measure of engagement, we did not differentiate between participant-generated and provider-generated messages. Participant messages may indicate engagement and self-motivation, while provider messages likely focus on feedback and support. This lack of differentiation limits the ability to fully interpret the specific roles and contributions of each message type in the observed outcomes. Additional factors such as the duration of interactions, the frequency of health-related queries, and the number of proactive messages sent can provide further context into how communication affects weight loss outcomes. This limitation may affect the accuracy of assessing the impact of communication on weight change rate. Thus, future studies should develop and use more comprehensive methods for evaluating communication, including qualitative assessments of message content and context. Furthermore, determining the effectiveness of different communication strategies, such as personalized feedback and motivational interviewing techniques, could provide insights into optimizing engagement and outcomes. Another limitation is that the proportion of days with weight recording in this study was calculated as the proportion of days on which participants recorded their weight over the entire 6-month period. While this measure provides an overall indicator of engagement, it does not capture temporal variations in engagement patterns, such as whether participants maintained consistent recording behavior throughout the intervention or showed high initial engagement followed by decline. This distinction is important because sustained engagement is often a critical factor for achieving successful behavior modification in mHealth interventions. Future study should explore longitudinal patterns of weight recording behavior to provide a more nuanced understanding of compliance and engagement trends in mHealth-based weight management programs.

Conclusions

The results of this study have revealed an association between weight change rate and the interaction of proportion of days with weight recording and communication volume between health care staff and participants in a lifestyle modification program. When the proportion of days with weight recording was high, the relation between communication volume and weight change rate was strong. Conversely, when the proportion of days with weight recording was low, the relation between communication volume and weight change rate was weak. These findings suggested that the acquisition of self-monitoring of health behaviors and lifelog data was crucial for maximizing the effectiveness of the health care professional's support. Furthermore, these insights would be beneficial in the development and process management of lifestyle modification programs using mHealth.

Acknowledgments

The authors thank the participants of the mobile app-based program for their contributions. Special thanks to the study team at PREVENT Inc. for their support in data collection and management. This work was supported by JSPS KAKENHI (grant 23K16769) and the Taiju Life Social Welfare Foundation.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

YH conceptualized the study, supervised data analysis, and drafted the manuscript. KS performed statistical analysis and contributed to manuscript preparation. SI performed statistical analysis and contributed to manuscript preparation. TA provided expertise on study design and critically revised the manuscript. MK assisted in data collection, manuscript writing, and revisions. TM provided oversight on methodology and reviewed the final manuscript.

Conflicts of Interest

This study was conducted as part of PREVENT, Inc's internal research activities. No external funding was received for this work. PREVENT, Inc provided access to anonymized program data and other necessary resources for the research. The company had no influence on the study design, data analysis, interpretation of findings, or the preparation of the manuscript, which were conducted independently by the authors.

Multimedia Appendix 1
Comparison between participants with complete data and excluded participants.
[DOCX File, 19 KB - i-jmr_v14i1e65863_app1.docx]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]

2. Okunogbe A, Nugent R, Spencer G, Powis J, Ralston J, Wilding J. Economic impacts of overweight and obesity: current and future estimates for 161 countries. BMJ Glob Health 2022 Sep;7(9):e009773. [doi: 10.1136/bmjgh-2022-009773] [Medline: 36130777]
3. Kotseva K, De Backer G, De Bacquer D, et al. Primary prevention efforts are poorly developed in people at high cardiovascular risk: a report from the European Society of Cardiology EURObservational Research Programme EUROASPIRE V survey in 16 European countries. Eur J Prev Cardiol 2021 May 8;28(4):370-379. [doi: 10.1177/2047487320908698] [Medline: 33966079]
4. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019 Dec;50(12):e344-e418. [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
5. Bairey Merz CN, Alberts MJ, Balady GJ, et al. ACCF/AHA/ACP 2009 competence and training statement: a curriculum on prevention of cardiovascular disease: a report of the American College of Cardiology Foundation/American Heart Association/American College of Physicians Task Force on Competence and Training (Writing Committee to Develop a Competence and Training Statement on Prevention of Cardiovascular Disease): developed in collaboration with the American Academy of Neurology; American Association of Cardiovascular and Pulmonary Rehabilitation; American College of Preventive Medicine; American College of Sports Medicine; American Diabetes Association; American Society of Hypertension; Association of Black Cardiologists; Centers for Disease Control and Prevention; National Heart, Lung, and Blood Institute; National Lipid Association; and Preventive Cardiovascular Nurses Association. Circulation 2009 Sep 29;120(13):e100-e126. [doi: 10.1161/CIRCULATIONAHA.109.192640] [Medline: 19770387]
6. Pearson GJ, Thanassoulis G, Anderson TJ, et al. 2021 Canadian Cardiovascular Society guidelines for the management of dyslipidemia for the prevention of cardiovascular disease in adults. Can J Cardiol 2021 Aug;37(8):1129-1150. [doi: 10.1016/j.cjca.2021.03.016] [Medline: 33781847]
7. Zomer E, Gurusamy K, Leach R, et al. Interventions that cause weight loss and the impact on cardiovascular risk factors: a systematic review and meta-analysis. Obes Rev 2016 Oct;17(10):1001-1011. [doi: 10.1111/obr.12433] [Medline: 27324830]
8. Dambha-Miller H, Day AJ, Strelitz J, Irving G, Griffin SJ. Behaviour change, weight loss and remission of Type 2 diabetes: a community-based prospective cohort study. Diabet Med 2020 Apr;37(4):681-688. [doi: 10.1111/dme.14122] [Medline: 31479535]
9. Jung J, Cho I. Promoting physical activity and weight loss with mhealth interventions among workers: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e30682. [doi: 10.2196/30682] [Medline: 35060913]
10. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 4;20(5):e162. [doi: 10.2196/jmir.9751] [Medline: 29728346]
11. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
12. Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A. The relationship between weight loss outcomes and engagement in a mobile behavioral change intervention: retrospective analysis. JMIR Mhealth Uhealth 2021 Nov 8;9(11):e30622. [doi: 10.2196/30622] [Medline: 34747706]
13. Zheng Y, Burke LE, Danford CA, Ewing LJ, Terry MA, Sereika SM. Patterns of self-weighing behavior and weight change in a weight loss trial. Int J Obes 2016 Sep;40(9):1392-1396. [doi: 10.1038/ijo.2016.68]
14. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mhealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
15. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
16. Zhaoyang R, Sliwinski MJ, Martire LM, Smyth JM. Social interactions and physical symptoms in daily life: quality matters for older adults, quantity matters for younger adults. Psychol Health 2019 Jul;34(7):867-885. [doi: 10.1080/08870446.2019.1579908] [Medline: 30821176]
17. Godino JG, Merchant G, Norman GJ, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755. [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
18. Little P, Stuart B, Hobbs FR, et al. Randomised controlled trial and economic analysis of an internet-based weight management programme: POWeR+ (Positive Online Weight Reduction). Health Technol Assess 2017 Jan;21(4):1-62. [doi: 10.3310/hta21040] [Medline: 28122658]
19. Arnold J, Bailey CP, Evans WD, Napolitano MA. Application of Mcguire's model to weight management messages: measuring persuasion of facebook posts in the healthy body, healthy u trial for young adults attending university in the

United States. Int J Environ Res Public Health 2022 Nov 1;19(21):14275. [doi: 10.3390/ijerph192114275] [Medline: 36361162]
20. Young AS, Cohen AN, Goldberg R, et al. Improving weight in people with serious mental illness: the effectiveness of computerized services with peer coaches. J Gen Intern Med 2017 Apr;32(Suppl 1):48-55. [doi: 10.1007/s11606-016-3963-0] [Medline: 28271427]
21. Kanai M, Toda T, Yamamoto K, Akimoto M, Hagiwara Y. A mobile health-based disease management program improves blood pressure in people with multiple lifestyle-related diseases at risk of developing vascular disease - a retrospective observational study. Circ Rep 2022 Jul 8;4(7):322-329. [doi: 10.1253/circrep.CR-22-0024] [Medline: 35860354]
22. Muramoto A, Matsushita M, Kato A, et al. Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes Res Clin Pract 2014;8(5):e466-e475. [doi: 10.1016/j.orcp.2013.10.003] [Medline: 25263836]
23. Edney S, Ryan JC, Olds T, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645. [doi: 10.2196/14645] [Medline: 31774402]
24. Berry R, Kassavou A, Sutton S. Does self-monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta-analysis. Obes Rev 2021 Oct;22(10):e13306. [doi: 10.1111/obr.13306] [Medline: 34192411]
25. Gonzalez C, Early J, Gordon-Dseagu V, Mata T, Nieto C. Promoting culturally tailored mHealth: a scoping review of mobile health interventions in latinx communities. J Immigr Minor Health 2021 Oct;23(5):1065-1077. [doi: 10.1007/s10903-021-01209-4] [Medline: 33988789]
26. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82. [doi: 10.2196/jmir.4100] [Medline: 27072817]
27. Lewis E, Hassmén P, Pumpa KL. Participant perspectives of a telehealth trial investigating the use of telephone and text message support in obesity management: a qualitative evaluation. BMC Health Serv Res 2021 Jul 9;21(1):675. [doi: 10.1186/s12913-021-06689-6] [Medline: 34243772]
28. Conroy MB, Yang K, Elci OU, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574. [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]
29. Frie K, Hartmann-Boyce J, Jebb S, Oke J, Aveyard P. Patterns in weight and physical activity tracking data preceding a stop in weight monitoring: observational analysis. J Med Internet Res 2020 Mar 17;22(3):e15790. [doi: 10.2196/15790] [Medline: 32181749]
30. Young HN, Len-Rios ME, Brown R, Moreno MM, Cox E. How does patient-provider communication influence adherence to asthma medications? Patient Educ Couns 2017 Apr;100(4):696-702. [doi: 10.1016/j.pec.2016.11.022] [Medline: 27916463]
31. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. Circulation 2019 Sep 10;140(11):e596-e646. [doi: 10.1161/CIR.0000000000000678] [Medline: 30879355]
32. Lahtio H, Rintala A, Immonen J, Sjögren T. The effectiveness of physical activity-promoting web- and mobile-based distance weight loss interventions on body composition in rehabilitation settings: systematic review, meta-analysis, and meta-regression analysis. J Med Internet Res 2022 Mar 24;24(3):e25906. [doi: 10.2196/25906] [Medline: 35323126]
33. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]

Abbreviations

mHealth: mobile health
NCD: noncommunicable disease

Edited by TDA Cardoso; submitted 28.08.24; peer-reviewed by G Jerome, M Yang; revised version received 15.02.25; accepted 18.02.25; published 27.03.25.

Please cite as:
Hagiwara Y, Adachi T, Kanai M, Shimizu K, Ishida S, Miki T
Interactive Effects of Weight Recording Frequency and the Volume of Chat Communication With Health Care Professionals on Weight Loss in mHealth Interventions for Noncommunicable Diseases: Retrospective Observational Study
Interact J Med Res 2025;14:e65863
URL: https://www.i-jmr.org/2025/1/e65863
doi:10.2196/65863

Review

Integration of Conventional and Virtual Reality Approaches in Augmented Reality for Theory-Based Psychoeducational Intervention Design for Chronic Low Back Pain: Scoping Review

Robin Conen[1], MSc; Steffen Mueller[2], Prof Dr; Ana Nanette Tibubos[1,3], Prof Dr

[1]Department of Nursing Science, Diagnostics in Healthcare and eHealth, Trier University, Trier, Germany

[2]Department of Computer Science/Therapeutic Science, Trier University of Applied Sciences, Trier, Germany

[3]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

Corresponding Author:
Robin Conen, MSc
Department of Nursing Science, Diagnostics in Healthcare and eHealth
Trier University
Max-Planck-Straße 6
Trier, 54296
Germany
Phone: 49 651 201 1904
Email: conen@uni-trier.de

Abstract

Background: Psychoeducation positively influences the psychological components of chronic low back pain (CLBP) in conventional treatments. The digitalization of health care has led to the discussion of virtual reality (VR) interventions. However, CLBP treatments in VR have some limitations due to full immersion. In comparison, augmented reality (AR) supplements the real world with virtual elements involving one's own body sensory perception and can combine conventional and VR approaches.

Objective: The aim of this study was to review the state of research on the treatment of CLBP through psychoeducation, including immersive technologies, and to formulate suggestions for psychoeducation in AR for CLBP.

Methods: A scoping review following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines was performed in August 2024 by using Livivo ZB MED, PubMed, Web of Science, American Psychological Association PsycINFO (PsycArticle), and PsyArXiv Preprints databases. A qualitative content analysis of the included studies was conducted based on 4 deductively extracted categories.

Results: We included 12 studies published between 2019 and 2024 referring to conventional and VR-based psychoeducation for CLBP treatment, but no study referred to AR. In these studies, educational programs were combined with physiotherapy, encompassing content on pain biology, psychological education, coping strategies, and relaxation techniques. The key outcomes were pain intensity, kinesiophobia, pain catastrophizing, degree of disability, quality of life, well-being, self-efficacy, depression, attrition rate, and user experience. Passive, active, and gamified strategies were used to promote intrinsic motivation from a psychological point of view. Regarding user experience from a software development perspective, user friendliness, operational support, and application challenges were recommended.

Conclusions: For the development of a framework for an AR-based psychoeducational intervention for CLBP, the combination of theories of acceptance and use of technologies with insights from health psychological behavior change theories appears to be of great importance. An example of a theory-based design of a psychoeducation intervention in AR for CLBP is proposed and discussed.

(Interact J Med Res 2025;14:e59611) doi:10.2196/59611

KEYWORDS

augmented reality; virtual reality; chronic low back pain; education; pain management; intervention

Introduction

Globally, 60%-80% of adults experience low back pain, with 10% developing chronic forms, of which 85% are classified as chronic nonspecific low back pain without a clear etiology [1]. Owing to the limited efficacy and adverse effects of pharmacological approaches, there is a need for nonpharmacological alternatives [2] to improve treatment outcomes [3] and develop effective behavioral interventions [4]. Treatment guidelines recommend behavioral modification, exercise, psychoeducation [5-7], and physiotherapy for trunk muscle strengthening [8-11] to reduce pain and disability.

Educational interventions for chronic low back pain (CLBP) provide knowledge about the condition, coping strategies, and physical activity [3,12], with the objective of enhancing the quality of life and symptom management by mitigating anxiety, kinesiophobia, hyperactive pain behavior, and depression, which are risk factors for pain chronification. Additionally, psychoeducation fosters self-efficacy to break the cycle between anxiety and pain [13,14].

Many traditional interventions to boost physical activity, which is key for CLBP treatment, rely on intention theories for modifying health behavior [15]. A prominent intention theory is the Unified Theory of Acceptance and Use of Technology 2 (UTAUT 2) by Venkatesh et al [16], which examines the acceptance and use of technologies. It has gained recognition in fields such as education, e-commerce, and health research with advancing health technology [17,18]. UTAUT 2 explains the formation of intention for technology use through the constructs of performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, and habit. These factors may also be useful for predicting the intentions of patients with CLBP toward educational technology adoption. Furthermore, insights from theories focusing on health behavior change may prove fruitful to consider when attempting to change health behavior through the use of a new technology. For instance, Schwarzer's Health Action Process Approach model [19], commonly used in health behavior interventions, highlights self-efficacy and outcome expectation. The Health Action Process Approach distinguishes between intention formation and implementation as well as between nonintenders, intenders, and actors, each requiring tailored interventions to promote self-efficacy, information, and support in implementation [20]. Another example is Michie's Behavior Change Technique (BCT) taxonomy with 93 BCTs outlining strategies for successful behavior change [21].

Immersive technologies can be characterized on the Reality-Virtuality Continuum by Milgram and Kishino [22]. They demonstrate visual display technologies ranging from real to virtual environments, including augmented reality (AR) and virtual reality (VR) [23-25]. AR enables the concurrent presence and interaction of digital and physical elements within real-world environments in real time. VR, in contrast, enables complete immersion in VR and represents the extreme of Milgram's continuum between reality and virtuality [26].

With regard to research in immersive technologies such as VR in the treatment of CLBP, VR-based treatments turned out to be promising in reducing acute, experimental, and chronic pain and can complement conventional CLBP treatments [27].

VR has proven effective in treating acute pain [24] by redirecting attention from unpleasant stimuli such as back pain to more pleasant visual, auditory, and tactile stimuli [27]. VR interventions were found to reduce pain intensity, catastrophizing symptoms, and psychological symptoms in patients with CLBP after one session through distraction, indicating the direct influence of VR on pain perception [26]. Other VR studies demonstrated the feasibility and efficacy of VR for CLBP as an alternative approach, such as VR applications with graded exposure during walking and grasping with integrated game design [28], self-administered VR therapy for CLBP at home [29], and its implementation even during COVID-19 [30]. A recent meta-analysis also showed that kinesiophobia and pain intensity in CLBP can be reduced through VR training [31].

Although there is some evidence for the safety and tolerability of VR treatment for CLBP, most studies lack methodological quality and results were limited to short-term effects. Studies on safety, acceptance, and satisfaction are lacking, including targeted investigations of the risks of spinal pain caused by VR [32]. Thus, while VR is promising in reducing CLBP symptoms, AR might offer additional benefits through the integration of physical and virtual elements, thereby reducing VR-associated discomfort. AR enables the coexistence and interaction of virtual and physical objects in real time in the real world, thus combining the advantages of VR while mitigating its limitations such as cybersickness and visual discomfort [33]. AR can enhance interaction, presence, intuitiveness, and pedagogical flexibility by enriching the real world with digital information, accommodating various learning styles, and facilitating teaching and learning [34]. Despite these presumed advantages of AR, to our knowledge, there are no empirical studies of AR-based treatment for CLBP.

In summary, pain treatment guidelines emphasize the key role of educational CLBP treatment to counteract psychological chronification and promote self-efficacy according to health behavior change models. Furthermore, when health behavior change is addressed using a new technology, a joint consideration of health psychological models with theories of acceptance from a technological perspective, like the UTAUT 2 [35], is considered useful for successful implementation. Existing studies with immersive technology [24,26-31] demonstrated positive effects for CLBP treatment in VR incorporating psychoeducational elements. However, these VR studies have methodological shortcomings and gaps regarding dimensions of user experience such as satisfaction and acceptance. Therefore, by formulating research questions using the PICO (Population, Intervention, Comparison, Outcome) framework, this scoping review aims to first examine research in patients with CLBP (P) receiving psychoeducation through immersive technology (I) compared to conventional psychoeducation (C) to improve pain relief and pain-psychological variables (O) and second on the basis of the results of the literature analysis, to develop an intervention design for AR-based psychoeducation in patients with CLBP that combines conventional methods with immersive technology

based on a technology acceptance model to promote acceptance and pain management.

Methods

We investigated the research question through a scoping review and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews [36]. This review includes studies that used psychoeducation for CLBP and chronic pain treatment: (1) conventionally, (2) with immersive technology in VR or AR, or (3) a combination of both, conventional therapy with VR or AR technology use. Only papers published in English or German in 2019-2024 were considered, wherein clinical guidelines were generally updated every 3-5 years with new evidence [37]. The exclusion criteria were as follows: (1) psychiatric patients, (2) acute back pain, (3) back pain after medical procedures, and (4) other specific pain conditions and pharmacological interventions. Scientific investigations or studies in journals or textbooks were included, regardless of the scientific methodology used. An electronic search was performed in August 2024 by using predefined English terms: ("chronic low back pain" OR "CLBP" OR "chronic pain") AND (("virtual reality" OR "augmented reality") OR ("education" OR "multimodal pain therapy" OR "psychological intervention")). Reviewer RC used Citavi to search for in vivo ZB MED and PubMed, and a manual search was conducted in the Web of Science, American Psychological Association PsycINFO, and PsyArXiv Preprints. The search was conducted in line with the Joanna Briggs Institute methodology for scoping reviews, extending the PRISMA statement [38]. In accordance with Arksey and O'Malley's [39] recommendations for scoping reviews, we did not include a formal quality assessment of the incorporated research. The selection process was initially based on a review of titles and abstracts regarding the inclusion and exclusion criteria, followed by an assessment of the full text by a reviewer (RC) and double-checked by another reviewer (ANT). Both reviewers (RC and ANT) extracted the following information from the included studies by using Microsoft Excel, following the Joanna Briggs Institute model: (1) citation, (2) context, (3) participant characteristics, (4) study aim, (5) methodology, (6) results, (7) interventions, (8) limitations, (9) key results related to review questions, and (10) future research areas [40]. Data analysis by the first reviewer (RC) utilized a qualitative content analysis [41]. A deductive approach was used to extract relevant categories for achieving the research objective. Guidelines for the treatment of CLBP [42] as well as recommendations of the World Health Organization for digital health interventions [43] served as the basis for this. Subsequently, 4 categories were extracted to capture all the essential aspects relevant to the design of the envisaged intervention. The categories are as follows: (1) content of CLBP-specific education, (2) factors based on the psychology of learning for the intervention design, (3) technical conditions (framework) for CLBP interventions, and (4) outcome measures of the educational interventions for CLBP.

Results

Study Selection

The study selection process, as shown in Figure 1, began with a database search that yielded 11,415 results. A total of 9602 titles were screened for the following terms: education, chronic pain, chronic back pain, CLBP, VR, and AR. The title should contain a minimum of 2 of the following keywords: education, chronic pain, chronic back pain, CLBP, VR, or AR; 9291 papers were excluded due to the lack of appearance of at least 2 of the defined terms. After applying the inclusion and exclusion criteria to 311 publications, 177 studies were excluded based on the titles and 95 were excluded based on the abstracts. After reviewing the full texts of the remaining 39 publications, 27 were excluded for (1) specific applications (eg, doctor-patient communication), (2) insufficient intervention descriptions, (3) overly specific populations (eg, elite athletes, primary school students, nursing staff), (4) unspecified psychoeducation (eg, cognitive behavioral therapy [CBT], cognitive functional training), and (5) unclear differentiation between education and physiotherapy in intervention design. Finally, the scoping review analyzed 12 publications, displayed in Table 1 [44-55].

The studies originated from Italy, Spain, the Netherlands, France, Chile, India, and Tunisia (8.33% each), Germany (25%), and the United States (16.67%). The review included 9 empirical studies (1 interview study, 8 interventional studies) and 3 reviews (1 systematic, 1 scoping, and 1 narrative review). All studies were peer-reviewed, except the narrative review. The results of the analysis of the 12 publications included are presented below according to the 4 defined categories.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the study selection process. APA: American Psychological Association; AR: augmented reality; CBP: chronic back pain; CLBP: chronic low back pain; VR: virtual reality.

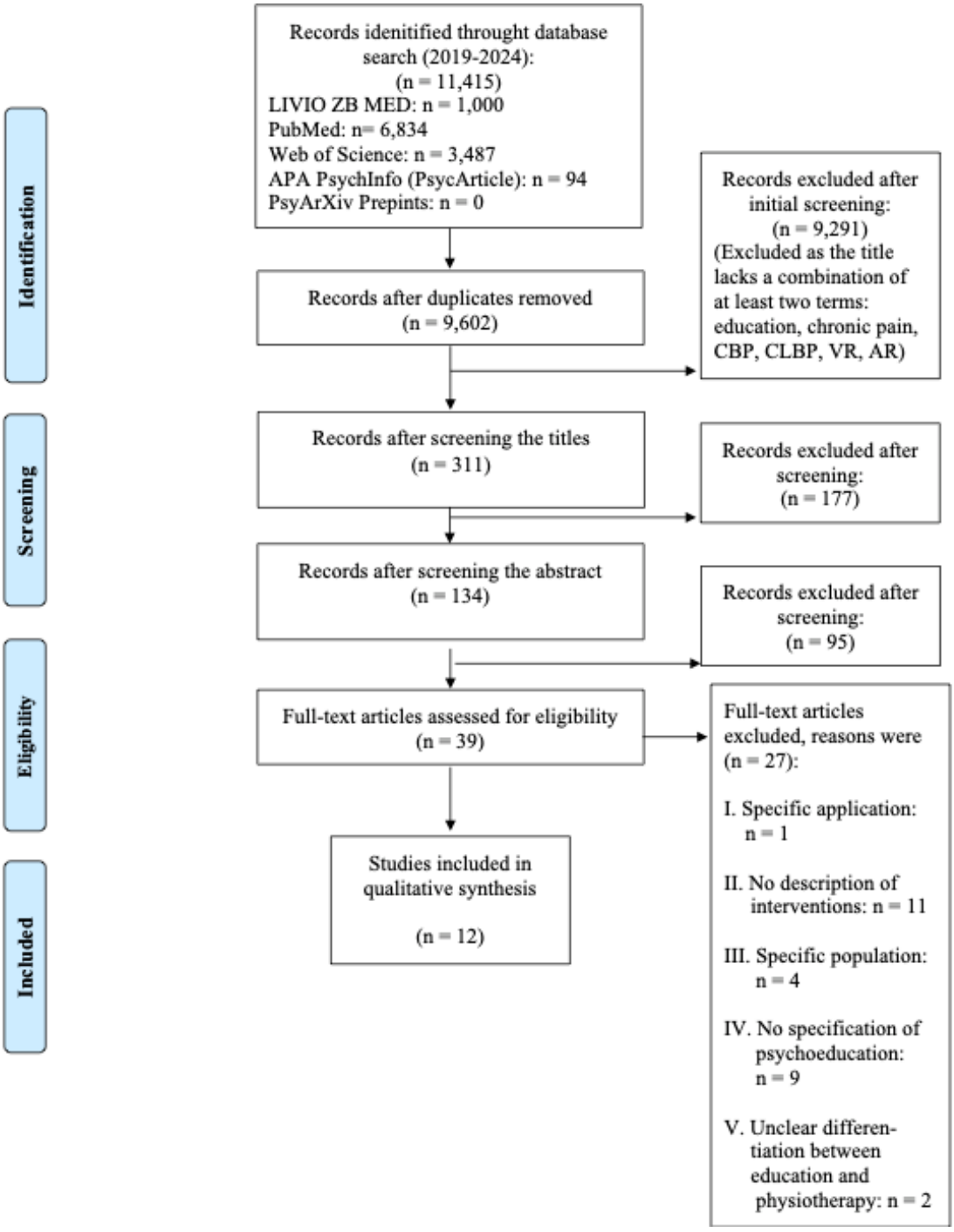


Table 1. Included studies applying psychoeducation by using conventional approaches and virtual reality approaches for the treatment of chronic low back pain [44-55].

Author	Study title	Journal name
Salazar-Méndez et al [44], 2024	Pain Neuroscience Education for Patients With Chronic Pain: A Scoping Review From Teaching-Leaning Strategies, Educational Level, and Cultural Perspective	Patient Education and Counseling
Ferlito et al [45], 2022	Pain Education in the Management of Patients with Chronic Low Back Pain: A Systematic Review	Journal of Functional Morphology and Kinesiology
Rim et al [46], 2022	Efficiency of Associating Therapeutic Patient Education with Rehabilitation in the Management of Chronic Low Back Pain: A Randomized Controlled Trial	Korean Journal of Family Medicine
Sidiq et al [47], 2024	Effects of Pain Education on Disability, Pain, Quality of Life, and Self-Efficacy in Chronic Low Back Pain: A Randomized Controlled Trial	PLOS One
Tomás-Rodríguez et al [48], 2024	Short- and Medium-Term Effects of a Single Session of Pain Neuroscience Education on Pain and Psychological Factors in Patients With Chronic Low Back Pain: A Single-Blind Randomized Clinical Trial	European Journal of Pain
Janik et al [49], 2024	Middle-Term Effects of Education Program in Chronic Low Back Pain Patients to an Adherence to Physical Activity: A Randomized Controlled Trial	Patient Education and Counseling
Lindner et al [50], 2020	Use of Virtual Reality as a Component of Acute and Chronic Pain Treatment	Anasthesiologie Intensivmedizin Notfallmedizin Schmerztherapie
Stamm et al [51], 2020	Virtual Reality in Pain Therapy: a Requirements Analysis for Older Adults With Chronic Back Pain	Journal of NeuroEngineering and Rehabilitation
Stamm et al [52], 2022	Virtual Reality Exergame for Supplementing Multimodal Pain Therapy in Older Adults With Chronic Back Pain	Virtual Reality
Brown et al [53], 2023	Chronic Pain Education Delivered With a Virtual Reality Headset in Outpatient Physical Therapy Clinics: A Multisite Exploratory Trial	American Journal of Translational Research
McConnell et al [54], 2024	A Multicenter Feasibility Randomized Controlled Trial Using a Virtual Reality Application of Pain Neuroscience Education for Adults With Chronic Low Back Pain	Annals of Medicine
de Vries et al [55], 2023	Pain Education and Pain Management Skills in Virtual Reality in the Treatment of Chronic Low Back Pain: A Multiple Baseline Single-Case Experimental Design	Behavior Research and Therapy

Contents of CLBP-Specific Education

The systematic review evaluated clinical studies from 2011 to 2021 comparing pain education/CBT with conventional physiotherapy for CLBP [45]. Thirteen studies, including 12 randomized controlled trials with 1642 participants, were analyzed. Six studies demonstrated a significant reduction in pain compared with the control group. The review concluded that due to the multimodality and heterogeneity of treatments, no definitive statement can be made regarding the efficacy of pain education or CBT in patients with CLBP [45]. Seven studies included an educational program in conjunction with physiotherapy [45-48,52-54]. Educational content varied considerably, ranging from exclusive focus on pain biology [46-49,52,53,55] to the inclusion of psychological aspects [46,52-55] and multidisciplinary approaches [49]. Most of the programs [48,55] incorporated education on pain physiology, frequently based on the book "Explain Pain" [56] by Butler and Moseley. Psychological education encompassed topics such as physical activity [45,46,49,53], fear of physical activity, emotional management [45-49,51], lifestyle modifications, daily exercises [45,49,52], pain-specific coping strategies [45,52,54], pain sensitization [47,54,55], and relaxation techniques, including stress management and mindfulness [45,49,52-55] in 6 studies, of which 5 included VR interventions. The content, duration, and physiotherapeutic integration of the individual education programs can be found in Multimedia Appendix 1.

Factors Based on Psychology of Learning for Intervention Design

This category encompasses factors of psychology of learning that are pertinent to the design of interactive interventions. Analysis of 3 studies showed that VR-based education employs passive mediation strategies such as informational videos and lectures (provided conventionally and in VR), alongside active and interactive strategies [44,50,54]. Three studies mentioned VR-based gamified approaches [50,52,55] and 2 studies [50,51] mentioned the promotion of intrinsic motivation.

Mediation Strategies

The included systematic review [44] examined the programs, cultural adaptations, and the efficacy of pain neuroscience education for chronic musculoskeletal pain, analyzing 71 studies that met our inclusion criteria and featured pain duration exceeding 3 months in adults. The analyzed studies explored pain neuroscience education in different settings by using

various experimental designs, including secondary analyses of randomized controlled trials, and showed positive effects on pain and psychological variables. Despite cultural influences on pain-relevant factors, only 2 (3%) of the 71 studies culturally adapted the pain neuroscience education material. Passive teaching-learning strategies tended to yield better outcomes for pain and functionality, whereas active methods resulted in significant knowledge improvements, albeit with insufficient description. The outcomes of multimodal therapies for chronic pain depend on the individualized integration of pain-specific education, considering biopsychosocial factors, educational level, culture, and diverse learning methods and materials for conveying pain neuroscience content [44]. Interaction content is passively conveyed through videos or lectures [46-49,53], particularly VR-based 360° nature videos [53].

Gamification and Motivation Enhancement

The Pain-Neuro-Education 2.0 software utilized a VR headset with immersive footage and computer-generated images for visually and emotionally engaging educational and relaxation training for chronic pain. This included interactive emotion regulation exercises such as breathing and mindfulness exercises in natural environments [54]. The VR program Recupt was also used to convey information in an engaging manner by having the user shoot at the pain stimulus with a laser gun, among other things. In the spinal cord phase, participants focus on visual “pain gates” and breathing to metaphorically “close” them and experience relaxation-induced pain relief. The brain component elucidates the reduction in pain response through the visualization and reactivation of illuminated connections. The alarm center gameplay demonstrates how emotions, cognitions, and behaviors influence pain perception. Finally, participants envision the alarm center as a brain region that regulates pain stimuli in an aircraft cockpit [55]. The VR program ViRST provides a therapeutic, interactive user interface with task-based activities in a farm environment [52]. Patients visualize movements and exertion levels by using game-based biofeedback with progress tracking and narrative elements [51]. Exergames incorporate biofeedback such as heart rate variability via photoplethysmography to prevent overexertion in interactive scenarios [52]. Gamification can motivate and enhance therapy adherence by fulfilling the psychological needs of competence, autonomy, and relatedness through interactive knowledge transfer. It also improves user skills through playful activities [50]. In the long term, feedback should be framed positively to maintain intrinsic motivation [52]. Avatars manipulate body perception for therapeutic effects, with the Proteus effect causing users to adopt their avatar’s behavior in real life. Personalized avatars can amplify pain relief [50].

Technical Conditions (Framework) for CLBP Intervention

The technical parameters of 3 enclosed VR studies, comprising 1 needs analysis [52] and 2 feasibility studies [52,53], provide insights into the design of AR-based education and identify potential areas of focus such as user-friendliness [51,53], operational support [51-53], and various application challenges [51,53]. The needs analysis was based on semistructured interviews (n=10) in focus groups to determine the requirements of older patients with chronic back pain, physiotherapists, and psychotherapists regarding VR pain therapy in terms of overall system, hardware, and software [51]. Findings emphasize that the designed system must be user-friendly; provide personalized instructions, demonstration videos, and individual guidance; and be available for rent. Assistants should support this system. Automatic breaks were considered crucial to avoid overexertion and pain aggravation. Activity should be limited to 30 minutes followed by a 15-minute rest. The study also highlighted the importance of balancing active therapy and relaxation. For hardware, it was determined that the VR headset must be independent and removable. Software design should consider user-friendliness by integrating the game environment with the level in-game environment for individual calibration of movement restrictions, particularly in gaming activities. Finally, a spacious room and wireless head-mounted display were considered essential for safety to prevent falls.

One feasibility study also emphasized the importance of safety aspects for the usability of VR headsets. The authors indicate that 93% of the application issues were associated with handling spatial and temporal limitations [53]. The second feasibility study demonstrated that disregarding body height (insufficient arm span) was perceived as disruptive [52]. Operational support software allows therapists to intervene during instances of pain, anxiety, or improper exercise execution by using a help button or emergency assistance [51]. Incorrect exercise execution is considered disruptive [52] and often lacks adequate support personnel for error correction or clinical assistance [53].

Outcome Measures of Educational Interventions for CLBP

Of the 12 studies [44-55] reviewed, 8 [46-49,54,55] were quantitative interventional studies. Commonly evaluated outcomes in CLBP studies encompassed pain intensity [45-49,51,52,54,55], kinesiophobia [46,48,51-53,55], pain catastrophizing [48,53-55], disability [45-47,51,52,54], health-related quality of life [51,52,54], well-being [47], self-efficacy [47,54], depression [46,53], attrition rate [49,53], and VR intervention user experience [52,53]. A comprehensive table displaying the different methods used, study results, and conclusions for each of the included studies can be found in Multimedia Appendix 2. A list of the pain-specific constructs and their measurement tools assessed in the different studies can be found in Multimedia Appendix 3.

Discussion

Principal Findings

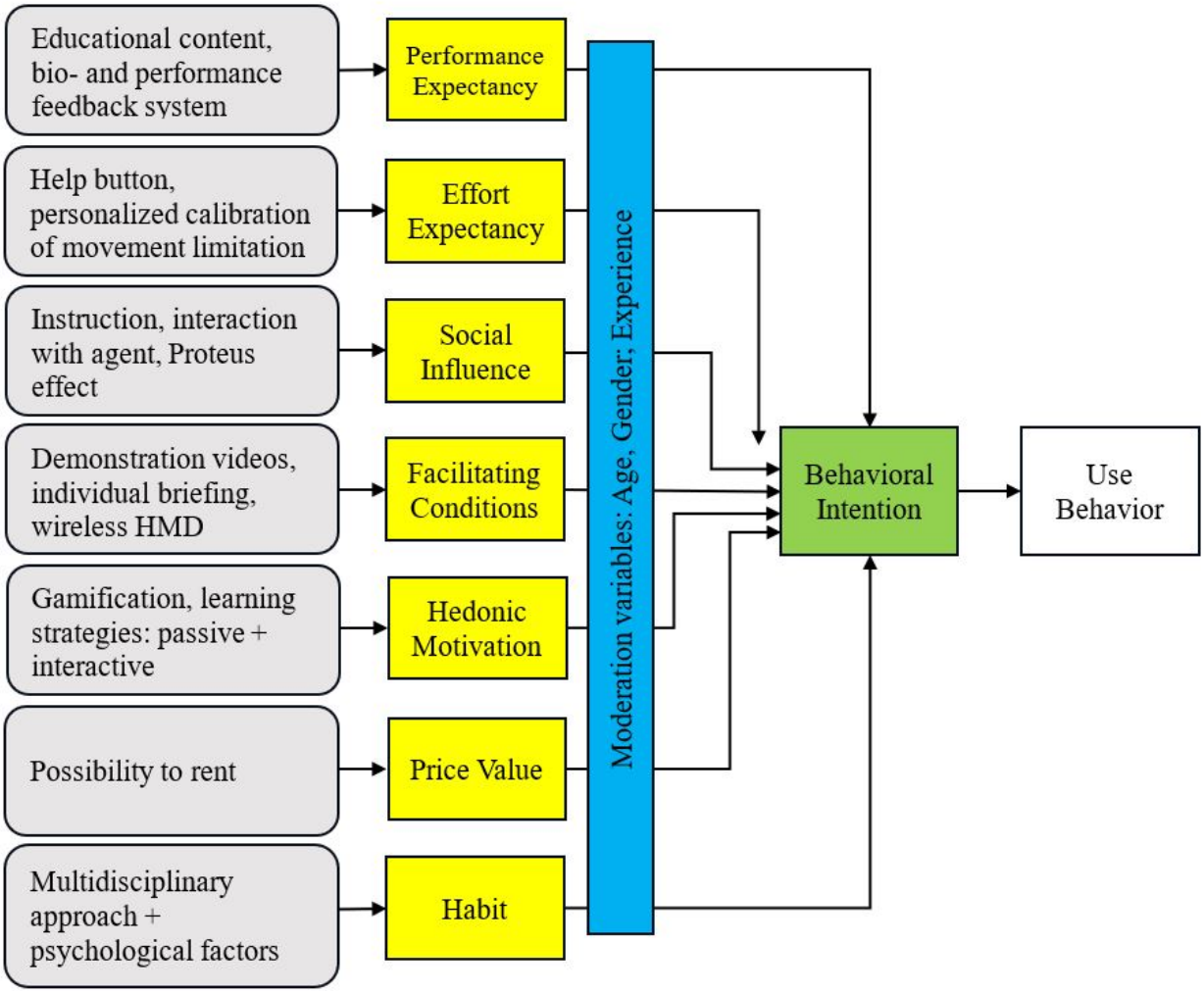
Psychoeducation is a key element for CLBP treatment. Psychoeducation provided in AR could offer more benefits than that in VR or with conventional methods by integrating physical and virtual elements. Additionally, psychoeducation in AR was assumed to be superior to VR due to reported VR-associated discomforts in CLBP treatment, such as cybersickness and visual discomfort. Therefore, we conducted a literature review in the first step to evaluate research on CLBP treatment through psychoeducation using conventional methods and immersive technologies in order to design a psychoeducational intervention

in AR for CLBP. In the second step, we applied the extracted results of the literature review to a theoretical framework, in particular, the UTAUT 2, to provide a design example for AR-based psychoeducation for CLBP.

Our findings indicate that various educational programs were combined with physiotherapy [45-48,52-54]. These studies referred to conventional methods or VR-based interventions. No relevant study with AR for CLBP treatment was found. The varying educational content encompassed pain biology [46-49], psychological education on physical activity [45,46,49,53], anxiety management [45-49,51], lifestyle modifications, daily exercises [45,49,52], coping strategies [45,52,54], pain sensitization [47,54,55], and relaxation techniques [45,49,52-55]. Passive, active [44,50,54], and gamified strategies [50,52,55] were employed alongside the promotion of intrinsic motivation [50,51]. User-friendliness [51,53], operational support [51,52,54], and application challenges [52,53] were considered important for software development. The key variables of educational CLBP interventions included physiological variables such as pain intensity [45-49,51,52,54,55] and disability level [45-47,51,52,54]; psychological variables such as kinesiophobia [46,48,51,52,54,55], pain catastrophizing [48,53-55], quality of life [51,52,55], well-being [47], self-efficacy [47,54], and depression [46,53]; and technical variables such as dropout rates [49,53] and user experience [52,53].

Our results elucidate key aspects of a useful design of a psychoeducational treatment in AR for CLBP, which does not exist to date, to the best of our knowledge. Our findings point out the relevance of the interplay of technical and psychological components, in particular, the health psychological aspects incorporating psychology of learning to foster behavior change. In the next step, the findings were applied to a theoretical framework. For this, we referred to the UTAUT 2 [8,9]. UTAUT 2 encompasses constructs such as (1) performance expectancy, (2) effort expectancy, (3) social influence, (4) facilitating conditions, (5) hedonic motivation, (6) price value, and (7) habit for intention formation as well as the moderating variables age, gender, and experience [8]. Therefore, we recommend the following design suggestions for psychoeducational interventions in AR based on UTAUT 2 for the treatment of CLBP, as exemplified in Figure 2.

Figure 2. Exemplary mapping of the extracted findings from the literature review (grey) applied to the Unified Theory of Acceptance and Use of Technology to design an artificial reality–based psychoeducation for chronic low back pain. HMD: head-mounted display.



UTAUT 2 Constructs

Performance Expectancy

It is recommended to convey psychoeducational content that demonstrates how CLBP can be positively influenced through physical activity [45,46,49,53], emotion management in kinesiophobia [45-49,51], pain-specific coping strategies [45,52,54], and stress and mindfulness techniques [45,49,52-55]. This content can be conveyed through a biofeedback system and supportive agent. As many educational measures are combined with physiotherapy [45-48,52-54], the agent can provide movement exercises, educational content, and interactive stress management techniques [54], supplemented by biofeedback. The biofeedback level and multimodal feedback of avatars promote top-down and bottom-up processes and enable associative learning [50]. To avoid discrepancies between instructions and sensory feedback, facilitate rapid corrections, and enhance user-friendliness [52], the avatar should provide immediate visual-acoustic performance feedback [50]. Biofeedback and body feedback are essential interventions for behavioral modification [17]. It is recommended to combine psychoeducation with a mindfulness-based stress reduction body scan and heart rate biofeedback [51], wherein biofeedback demonstrates progress and enables gamification elements [50,52].

Effort Expectancy

Software design should incorporate a gaming environment with in-game level settings to facilitate personalized calibration of movement limitations and body size, particularly for therapeutic activities [51,52]. The system should enable therapists to intervene through a help button or emergency assistance when patients experience pain, anxiety, or perform exercises incorrectly [51]. Frequent feedback for incorrect exercise execution should be avoided, as it may be perceived as disruptive [52].

Social Influence

With regard to social influence, an AR intervention should be accompanied by an agent that conveys pain-specific knowledge through lectures as passive knowledge transfer [44] or through support in psychological interactions, such as stress management exercises [54]. An agent can monitor a patient's body movements to integrate the phenomenon of "virtual body ownership" into the body image or utilize the analgesic effect of the Proteus effect to promote behaviors in the real world [50]. This aligns with BCT's recommendations for behavior change, wherein instruction, repetition, and demonstration of behavior have positive effects on physical activity that persist for up to 6 months [50,51]. Therefore, we propose to increase the intention to use by incorporating an agent in an AR intervention, with both passive and interactive roles.

Facilitating Conditions

A user-friendly system requires personalized instructions, demonstration videos, and individual briefings. For safety considerations, a spacious environment and wireless head-mounted displays are essential to mitigate the risk of falls. An assistant should be present to support the system. Automated breaks are crucial to prevent overexertion and exacerbation of pain, thereby automatically balancing the active therapy and relaxation periods. The VR headset should be designed for independent removal and application [51].

Hedonic Motivation

Passive learning strategies tend to yield superior outcomes in pain and functionality, whereas active methods can elicit significant improvements in knowledge [44]. Gamification demonstrates a motivation-enhancing effect through active and interactive patient engagement [50,52], for instance, through the interactive development of pain-specific knowledge in the Reducept program, where users enter their own brains and shoot with laser guns or connect points [55], or through movement exercises on a simulated farm [52]. A feedback system or biofeedback could be integrated, as outlined in the variable "performance expectancy" of UTAUT 2 and should be phrased positively as praise to increase motivation [51]. Praise as a social reward can occur through interaction with an agent, as described in the variable "social influence" [18]. In CBT, praise serves as positive reinforcement to promote adaptive behaviors and cognitions corresponding to positive CBT, which incorporates positive psychology and solution-focused brief therapy into a cognitive-behavioral context [57]. Gamification in an AR-based intervention enables the implementation of BCTs [17] by creating a material incentive such as within the framework of a game in an AR application [43].

Price Value

The headsets required for the interventions should be provided or loaned rather than purchased [50].

Habit

Educational programs for CLBP should incorporate a multidisciplinary approach that encompasses both physiological and psychological pain while promoting behavioral modifications such as regular physical activity [49]. Conventional recommendations for behavioral change emphasize repetition as crucial for habit formation [50]. For long-term interventions aimed at behavioral modification, theories addressing the intention-behavior gap and behavioral automaticity in physical activity should be considered, such as the Affective Reflective Theory of Physical Inactivity and Exercise [58] or the Physical Activity Adoption and Maintenance model [59].

Strengths and Limitations

This scoping review gives an overview of the most important educational content, elements of psychological training, interactive design forms, and relevant pain psychological variables for developing CLBP interventions in AR. It offers a substantiated basis for a theory-based development of a psychoeducational treatment in AR. Thus, this study provides a framework for the theory-driven extraction of hypotheses for future AR research in CLBP treatment. One limitation of this review encompasses the exclusion of certain sport science and physiotherapy databases (eg, SPORTDiscus) and the restriction to studies published in German and English, potentially omitting relevant publications. Further, the distinction between CLBP and chronic nonspecific low back pain in the included studies

was often imprecise. The distinction might be relevant for the intervention design, which was neglected in our analysis.

Recommendations for Research

First, the theoretically proposed design of AR-based psychoeducation for CLBP should be realized in future research. Second, an evaluation of the feasibility and user experience is needed. Third, the therapeutic efficacy of the psychoeducational content in AR must be demonstrated in a clinical evaluation study with patients with CLBP. As no studies on the psychometric properties of measurements in AR are known, psychometric assessments must be tested for measurement equivalence.

Conclusions

For the development of a framework for an AR-based psychoeducational intervention in CLBP, the combination of theories of acceptance and use of technologies with insights from health psychological behavior change theories appears to be of great importance. An example for a theory-based design of psychoeducation in AR for CLBP is proposed and discussed. Our results offer a substantiated basis for a theory-based development of psychoeducational treatment in AR.

Acknowledgments

This work was funded in part by the Ministry for Science and Health of Rhineland-Palatinate as part of the research training group (Forschungskolleg) “Immersive Extended Reality for Physical Activity and Health” (XR-PATH). The publication was funded or supported by the Open Access Fund of Universität Trier and by the German Research Foundation (DFG).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Contents of the education, duration, and physiotherapy involvement.
[DOCX File , 18 KB - ijmr_v14i1e59611_app1.docx]

Multimedia Appendix 2
Summary of the study characteristics, results, and conclusions of the 12 included studies.
[DOCX File , 19 KB - ijmr_v14i1e59611_app2.docx]

Multimedia Appendix 3
Summary of the pain-specific measuring outcomes and instruments.
[DOCX File , 17 KB - ijmr_v14i1e59611_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews checklist.
[DOCX File , 108 KB - ijmr_v14i1e59611_app4.docx]

References

1. Alsufiany M, Lohman E, Daher N, Gang G, Shallan A, Jaber H. Non-specific chronic low back pain and physical activity: A comparison of postural control and hip muscle isometric strength: A cross-sectional study. Medicine (Baltimore) 2020 Jan;99(5):e18544 [FREE Full text] [doi: 10.1097/MD.0000000000018544] [Medline: 32000363]
2. Schreijenberg M, Koes BW, Lin CC. Guideline recommendations on the pharmacological management of non-specific low back pain in primary care - is there a need to change? Expert Rev Clin Pharmacol 2019 Feb;12(2):145-157. [doi: 10.1080/17512433.2019.1565992] [Medline: 30618319]
3. Faucett J. Chronic low back pain: early interventions. Annu Rev Nurs Res 1999 Jan 01;17(1):155-182. [doi: 10.1891/0739-6686.17.1.155]
4. Dobbels F, De Bleser L, Berben L, Kristanto P, Dupont L, Nevens F, et al. Efficacy of a medication adherence enhancing intervention in transplantation: The MAESTRO-Tx trial. J Heart Lung Transplant 2017 May;36(5):499-508. [doi: 10.1016/j.healun.2017.01.007] [Medline: 28162931]
5. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, COST B13 Working Group on Guidelines for Chronic Low Back Pain. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15 Suppl 2(Suppl 2):S192-S300 [FREE Full text] [doi: 10.1007/s00586-006-1072-1] [Medline: 16550448]
6. Schäfer R, Niederer D, Levenig C, et al. Effectiveness and mechanisms of a multimodal treatment for low back pain: A pragmatic matched cohort study. MedRxiv. Preprint posted online on February 2, 2023 2023. [doi: 10.1101/2022.09.28.22280380]

7. Alfuth M, Lages A. Chronischer Rückenschmerz. CME 2018 Jun 13;15(6):9-24. [doi: 10.1007/s11298-018-6424-8]
8. Frizziero A, Pellizzon G, Vittadini F, Bigliardi D, Costantino C. Efficacy of core stability in non-specific chronic low back pain. JFMK 2021 Apr 22;6(2):37. [doi: 10.3390/jfmk6020037]
9. Mueller J, Niederer D. Dose-response-relationship of stabilisation exercises in patients with chronic non-specific low back pain: a systematic review with meta-regression. Sci Rep 2020 Oct 09;10(1):16921 [FREE Full text] [doi: 10.1038/s41598-020-73954-9] [Medline: 33037280]
10. Minobes-Molina E, Nogués MR, Giralt M, Casajuana C, de Souza DLB, Jerez-Roig J, et al. Effectiveness of specific stabilization exercise compared with traditional trunk exercise in women with non-specific low back pain: a pilot randomized controlled trial. PeerJ 2020;8:e10304 [FREE Full text] [doi: 10.7717/peerj.10304] [Medline: 33312766]
11. Sutanto D, Ho RST, Poon ETC, Yang Y, Wong SHS. Effects of different trunk training methods for chronic low back pain: a meta-analysis. IJERPH 2022 Mar 01;19(5):2863. [doi: 10.3390/ijerph19052863]
12. Isaikin AI, Nasonova TI, Mukhametzyanova AK. Emotional disorders and their therapy in chronic low back pain. RJTAO 2022 Oct 25;14(5):90-95. [doi: 10.14412/2074-2711-2022-5-90-95]
13. Comachio J, Magalhães MO, Campos Carvalho E Silva APDM, Marques AP. A cross-sectional study of associations between kinesiophobia, pain, disability, and quality of life in patients with chronic low back pain. Adv Rheumatol 2018 Jun 22;58(1):8. [doi: 10.1186/s42358-018-0011-2] [Medline: 30657061]
14. Williams FMK, Elgaeva EE, Freidin MB, Zaytseva OO, Aulchenko YS, Tsepilov YA, et al. Causal effects of psychosocial factors on chronic back pain: a bidirectional Mendelian randomisation study. Eur Spine J 2022 Jul;31(7):1906-1915 [FREE Full text] [doi: 10.1007/s00586-022-07263-2] [Medline: 35662366]
15. Silva-Smith AL, Hanson CL, Neubeck L, Rowat A, McHale S. Physical activity interventions framed by the health action process approach for adults with long-term conditions: a scoping review. Int J Behav Med 2024 Dec;31(6):987-1017. [doi: 10.1007/s12529-024-10305-2] [Medline: 39009797]
16. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
17. Or C. Towards an integrated model: task-technology fit in unified theory of acceptance and use of technology 2 in education contexts. JALT 2023 Jan 20;6(1):1-2. [doi: 10.37074/jalt.2023.6.1.8]
18. Barbosa RAP, Martins AVS, Da Silva IPA, Costa LA, Oliveira RA, De Souza HC. Acceptance and use of a virtual learning environment (VLE): structural equations modeling of the unified theory of acceptance and use of technology. Int J Innov Educ Res 2020 Apr 01;8(4):237-244. [doi: 10.31686/ijier.vol8.iss4.2285]
19. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
20. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. Correction: A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2019 Feb 06;21(2):e12015 [FREE Full text] [doi: 10.2196/12015] [Medline: 30726188]
21. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst. 1994. URL: https://cs.gmu.edu/~zduric/cs499/Readings/r76JBo-Milgram_IEICE_1994.pdf [accessed 2024-01-13]
23. Skarbez R, Smith M, Whitton MC. Revisiting Milgram and Kishino's reality-virtuality continuum. Front Virtual Real 2021 Mar 24;2:1. [doi: 10.3389/frvir.2021.647997]
24. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019;12:2053-2085 [FREE Full text] [doi: 10.2147/JPR.S200498] [Medline: 31308733]
25. Awan A. Enhancing user experience: exploring mobile augmented reality experiences. VFAST Transactions on Software Engineering 2024:121-132. [doi: 10.21015/vtse.v12i1.1741]
26. Trujillo MS, Alvarez AF, Nguyen L, Petros J. Embodiment in virtual reality for the treatment of chronic low back pain: a case series. J Pain Res 2020;13:3131-3137 [FREE Full text] [doi: 10.2147/JPR.S275312] [Medline: 33269003]
27. Tack C. Virtual reality and chronic low back pain. Disabil Rehabil Assist Technol 2021 Aug;16(6):637-645. [doi: 10.1080/17483107.2019.1688399] [Medline: 31746250]
28. Hennessy RW, Rumble D, Christian M, Brown DA, Trost Z. A graded exposure, locomotion-enabled virtual reality app during walking and reaching for individuals with chronic low back pain: cohort gaming design. JMIR Serious Games 2020 Aug 10;8(3):e17799 [FREE Full text] [doi: 10.2196/17799] [Medline: 32773381]
29. Garcia LM, Darnall BD, Krishnamurthy P, Mackey IG, Sackman J, Louis RG, et al. Correction: self-administered behavioral skills-based at-home virtual reality therapy for chronic low back pain: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Feb 12;10(2):e27652 [FREE Full text] [doi: 10.2196/27652] [Medline: 33577465]
30. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled

trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
31. Li R, Li Y, Kong Y, Li H, Hu D, Fu C, et al. Virtual reality-based training in chronic low back pain: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2024 Feb 26;26:e45406 [FREE Full text] [doi: 10.2196/45406] [Medline: 38407948]
32. Nagpal AS, Raghunandan A, Tata F, Kibler D, McGeary D. Virtual reality in the management of chronic low back pain: a scoping review. Front Pain Res (Lausanne) 2022;3:856935 [FREE Full text] [doi: 10.3389/fpain.2022.856935] [Medline: 35295809]
33. Wu YF, Kim EY. Users' perceptions of technological features in augmented reality (AR) and virtual reality (VR) in fashion retailing: a qualitative content analysis. Mobile Information Systems 2022 Aug 4;2022:1-13. [doi: 10.1155/2022/3080280]
34. Sala N. IGI Global. Virtual reality, augmented reality, and mixed reality in education 2020:48-73. [doi: 10.4018/978-1-7998-4960-5.ch003]
35. Tamilmani K, Rana NP, Wamba SF, Dwivedi R. The extended Unified Theory of Acceptance and Use of Technology (UTAUT2): A systematic literature review and theory evaluation. International Journal of Information Management 2021 Apr;57:102269. [doi: 10.1016/j.ijinfomgt.2020.102269]
36. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
37. Vernooij RW, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. Implement Sci 2014 Jan 02;9:3 [FREE Full text] [doi: 10.1186/1748-5908-9-3] [Medline: 24383701]
38. von Elm E, Schreiber G, Haupt CC. [Not Available]. Z Evid Fortbild Qual Gesundhwes 2019 Jun;143:1-7. [doi: 10.1016/j.zefq.2019.05.004] [Medline: 31296451]
39. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616] [Medline: 21410846]
40. Khalil H, Peters MD, Tricco AC, Pollock D, Alexander L, McInerney P, et al. Conducting high quality scoping reviews-challenges and solutions. J Clin Epidemiol 2021 Feb;130:156-160. [doi: 10.1016/j.jclinepi.2020.10.009] [Medline: 33122034]
41. Mayring P, Fenzl T. Qualitative Inhaltsanalyse. In: Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer; 2019.
42. Nicol V, Verdaguer C, Daste C, Bisseriex H, Lefèvre-Colau MM, Rannou F, et al. Chronic low back pain: a narrative review of recent international guidelines for diagnosis and conservative treatment. J Clin Med 2023 Feb 20;12(4):1685 [FREE Full text] [doi: 10.3390/jcm12041685] [Medline: 36836220]
43. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019.
44. Salazar-Méndez J, Cuyul-Vásquez I, Ponce-Fuentes F, Guzmán-Muñoz E, Núñez-Cortés R, Huysmans E, et al. Pain neuroscience education for patients with chronic pain: A scoping review from teaching-learning strategies, educational level, and cultural perspective. Patient Educ Couns 2024 Jun;123:108201. [doi: 10.1016/j.pec.2024.108201] [Medline: 38387389]
45. Ferlito R, Blatti C, Lucenti L, Boscarino U, Sapienza M, Pavone V, et al. Pain education in the management of patients with chronic low back pain: a systematic review. JFMK 2022 Sep 26;7(4):74. [doi: 10.3390/jfmk7040074]
46. Rim M, Leila R, Aicha BT, Olfa S, Meriem H, Ines L, et al. Efficiency of associating therapeutic patient education with rehabilitation in the management of chronic low back pain: a randomized controlled trial. Korean J Fam Med 2022 Nov;43(6):367-373 [FREE Full text] [doi: 10.4082/kjfm.21.0223] [Medline: 36444121]
47. Sidiq M, Muzaffar T, Janakiraman B, Masoodi S, Vasanthi RK, Ramachandran A, et al. Effects of pain education on disability, pain, quality of life, and self-efficacy in chronic low back pain: A randomized controlled trial. PLoS One 2024;19(5):e0294302 [FREE Full text] [doi: 10.1371/journal.pone.0294302] [Medline: 38805446]
48. Tomás-Rodríguez MI, Asensio-García MDR, García RN, Delicado-Miralles M, Sánchez SH, Segura-Heras JV. Short- and medium-term effects of a single session of pain neuroscience education on pain and psychological factors in patients with chronic low back pain. A single-blind randomized clinical trial. Eur J Pain 2024 Nov;28(10):1841-1854. [doi: 10.1002/ejp.4700] [Medline: 39017623]
49. Janik F, Fabre C, Seichepine AL, Masquelier B, Barbier F, Toulotte C. Middle-term effects of education programme in chronic low back pain patients to an adherence to physical activity: A randomized controlled trial. Patient Education and Counseling 2024 Feb;119:108081. [doi: 10.1016/j.pec.2023.108081]
50. Lindner S, Latoschik M, Rittner H. [Use of virtual reality as a component of acute and chronic pain treatment]. Anasthesiol Intensivmed Notfallmed Schmerzther 2020 Sep;55(9):549-561. [doi: 10.1055/a-1022-3038] [Medline: 32916738]
51. Stamm O, Dahms R, Müller-Werdan U. Virtual reality in pain therapy: a requirements analysis for older adults with chronic back pain. J Neuroeng Rehabil 2020 Sep 29;17(1):129 [FREE Full text] [doi: 10.1186/s12984-020-00753-8] [Medline: 32993678]

52. Stamm O, Dahms R, Reithinger N, Ruß A, Müller-Werdan U. Virtual reality exergame for supplementing multimodal pain therapy in older adults with chronic back pain: a randomized controlled pilot study. Virtual Real 2022;26(4):1291-1305 [FREE Full text] [doi: 10.1007/s10055-022-00629-3] [Medline: 35194374]
53. Brown L, DiCenso-Fleming T, Ensign T, Boyd AJ, Monaghan G, Binder DS. Chronic pain education delivered with a virtual reality headset in outpatient physical therapy clinics: a multi-site exploratory trial. Am J Transl Res 2023;15(5):3500-3510 [FREE Full text] [Medline: 37303618]
54. McConnell R, Lane E, Webb G, LaPeze D, Grillo H, Fritz J. A multicenter feasibility randomized controlled trial using a virtual reality application of pain neuroscience education for adults with chronic low back pain. Ann Med 2024 Dec;56(1):2311846 [FREE Full text] [doi: 10.1080/07853890.2024.2311846] [Medline: 38354690]
55. de Vries FS, van Dongen RTM, Bertens D. Pain education and pain management skills in virtual reality in the treatment of chronic low back pain: A multiple baseline single-case experimental design. Behav Res Ther 2023 Mar;162:104257. [doi: 10.1016/j.brat.2023.104257] [Medline: 36731183]
56. Butler D, Moseley G. Explain pain. In: 2nd ed. Australia: Noigroup Publications; 2013.
57. Bannink F, Geschwind N. Positive CBT: Individual and Group Treatment Protocols for Positive Cognitive Behavioral Therapy. Göttingen, Germany: Hogrefe; May 2021.
58. Brand R, Ekkekakis P. Affective–Reflective Theory of physical inactivity and exercise. Ger J Exerc Sport Res 2017 Nov 14;48(1):48-58. [doi: 10.1007/s12662-017-0477-9]
59. Jekauc D, Gürdere C, Englert C, Strobach T, Bottesi G, Bray S, et al. The contribution and interplay of implicit and explicit processes on physical activity behavior: empirical testing of the physical activity adoption and maintenance (PAAM) model. BMC Public Health 2024 May 06;24(1):1239 [FREE Full text] [doi: 10.1186/s12889-024-18589-5] [Medline: 38711051]

Abbreviations

AR: augmented reality
BCT: behavior change technique
CBT: cognitive behavioral therapy
CLBP: chronic low back pain
PICO: Population, Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UTAUT 2: Unified Theory of Acceptance and Use of Technology 2
VR: virtual reality

Edited by M Yousufuddin; submitted 17.04.24; peer-reviewed by A Werner, Z Lu, L Pérez-Gisbert; comments to author 30.07.24; revised version received 07.11.24; accepted 10.11.24; published 20.01.25.

Please cite as:
Conen R, Mueller S, Tibubos AN
Integration of Conventional and Virtual Reality Approaches in Augmented Reality for Theory-Based Psychoeducational Intervention Design for Chronic Low Back Pain: Scoping Review
Interact J Med Res 2025;14:e59611
URL: https://www.i-jmr.org/2025/1/e59611
doi:10.2196/59611
PMID:

Original Paper

Simplified Medication Adherence Questionnaire (SMAQ) for People Living With HIV in a National Hospital in Mexico: Instrument Validation Study

Luis Eduardo Del Moral Trinidad[1], MPH; Luz Alicia González Hernández[2], PhD; Jaime Federico Andrade Villanueva[2], MCR; Pedro Martínez-Ayala[2], MCR; Adriana Valle Rodríguez[2], MD; Vida Veronica Ruíz Herrera[2], MD; José Adán Vizcaíno Résendiz[2], MOP; Melva Guadalupe Herrera Godina[3], PhD; Sergio Dominguez-Lara[4], PhD

[1]Department of Public Health Sciences, University of Guadalajara, Guadalajara, Mexico

[2]HIV Unit, Hospital Civil de Guadalajara, Guadalajara, Mexico

[3]Institute of Biomedical Sciences Research, University of Guadalajara, Guadalajara, Mexico

[4]Research Institute, Facultad de Ciencias de la Comunicación, Turismo y Psicología, University of San Martín de Porres, Lima, Peru

Corresponding Author:
Sergio Dominguez-Lara, PhD
Research Institute
Facultad de Ciencias de la Comunicación, Turismo y Psicología
University of San Martín de Porres
Avenue Tomás Marsano 242 (5th floor)
Lima, 15048
Peru
Phone: 51 5136300
Email: sdominguezmpcs@gmail.com

Abstract

Background: Adherence to antiretroviral therapy is a critical component in achieving viral suppression in people living with HIV in addition to increasing overall quality of life. Several indirect methods have been used to measure adherence including the Simplified Medication Adherence Questionnaire (SMAQ).

Objective: The objective of this study is to evaluate the reliability and validity of the SMAQ in men living with HIV/AIDS attending a Mexican national hospital.

Methods: A cross-sectional analytical design study was carried out in a Mexican National Hospital in Jalisco, including men aged >18 years with at least 3 months of antiretroviral treatment, excluding those with cognitive difficulties in answering the survey. A minimum sample size was calculated to detect the contribution of the variables within the model. The analysis included descriptive tests, confirmatory factor analysis, reliability and validity assessment, correlation between adherence and viral load, and association between viral load and adherence.

Results: The final analysis included a total of 260 patients with a mean age of 43 (SD 12) years and an average of 8.97 (SD 6.33) years on antiretroviral treatment. The SMAQ showed sufficient structural validity (comparative fit index=1, root-mean-square error of approximation=0, 90% CI 0-0.085) with satisfactory factor loadings on most questions except item 2 (Do you always take your medication at the prescribed time?). The reliability of the scale is acceptable (Cronbach α=0.702, ω=0.718). Adherence correlated with viral load significantly but not with recent TCD4 lymphocyte levels. Patients classified as adherent were three times more likely to be undetectable than nonadherent patients (odds ratio 3.31, 95% CI 1.13-9.64, *P*=.04).

Conclusions: The SMAQ represents an adequate tool to assess adherence in men living with HIV in the Mexican context, this will contribute to this study and compression of adherence to establish future intervention programs.

(Interact J Med Res 2025;14:e59562) doi:10.2196/59562

KEYWORDS

treatment adherence; HIV; Mexico; validation; Spanish; Hispanic; cross sectional; surveys; questionnaires; scales; adherence; viral load; sexually transmitted infection; STI; drugs; pharmacotherapy; medication; simplified medication adherence questionnaire; SMAQ

Introduction

According to the Joint United Nations Program on HIV/AIDS, 39 million people were living with HIV by 2022, of which only 76.4% were receiving antiretroviral treatment (ART). In Mexico, by 2022, there were 270,000 cases registered with the Ministry of Health, 80% (n=270,000) of which were men; in addition, Jalisco ranks 4th in prevalence of people living with HIV with a record of 7134 patients on ART and 78% of this population has achieved viral suppression through such treatment [1,2].

The World Health Organization defines adherence to treatment as an individual's behavior regarding medications, diet, and lifestyle changes that correspond to the recommendations provided by a health professional [3].

Therapeutic adherence is a complex process that is made up of a personal component represented by the patient, where their attitudes toward their disease and the positions they take on it are concentrated, as well as a relational component involving the health professional and the health structure that surrounds them. All these components work synergistically toward a common goal to benefit the patient's health [4].

Adherence to ART is a critical component to achieve viral suppression in people living with HIV as well as to increase the overall quality of life [5]. Its study is important because in Mexico there has been a sustained prevalence in recent years, and it is necessary to improve the tools available to optimize treatment success [6].

There are various methods for measuring adherence in a patient; however, there is no gold standard for this purpose. Due to this, its measurement will depend on the characteristics of the population studied; in addition, the method should have basic psychometric standards of acceptable validity and reliability [7]. These methods are classified as direct, those that directly quantify the drug and its metabolites in blood or any other fluid or tissue, but they are costly and impractical for routine implementation, and indirect, that is they base their measurement on pill counts and self-reports, among others [8].

An indirect tool that has been used by several authors is the Simplified Medication Adherence Questionnaire (SMAQ). This questionnaire has been used in studies to evaluate adherence to ART in people living with HIV, and its six-question structure makes it practical in clinical contexts where a rapid evaluation is required [9,10]. This scale was introduced in 2002 by Knobel et al [9]. They designed the scale intending to create a questionnaire to identify nonadherent patients and found that this instrument had a sensitivity of 72%, a specificity of 91%, and a likelihood ratio of 7.94, as well as a Cronbach α of 0.75 [9].

The SMAQ questionnaire has also been validated in different pathologies, medical conditions, and chronic diseases, such as hypertension, diabetes, and tuberculosis. These validations have shown consistent results, supporting the usefulness and validity of the questionnaire in different health care settings [9,11]. The instrument was originally developed for the Spanish population in 1999, but the current social and pharmacological context differs for the population we are now studying. It is important to evaluate the validity and reliability of the instrument when used in contexts different from its original development, to ensure the quality of the collected information. This is because the metric quality of a self-report questionnaire must be explored in the context where it will be applied. Otherwise, the psychometric properties would be compromised, leading to negative consequences for the evaluation [12,13].

This questionnaire is not yet validated in Latin people, thus, its utility in our population is unknown. There is only a pilot study with 10 participants that used the SMAQ in Peruvian people living with HIV; where item comprehension was evaluated, as cultural applicability and social acceptance [14]. To our knowledge, no studies have been conducted to validate this instrument in the Mexican population. This study aimed to evaluate the reliability and validity of the SMAQ in people living with HIV/AIDS attending a Mexican national hospital.

Methods

Study Design

This study is a cross-sectional analytical design study concerning people living with HIV/AIDS receiving ART at the Civil Hospital of Guadalajara "Fray Antonio Alcalde."

Selection of Participants

Men older than 18 years of age receiving ART for at least three months before their inclusion in this study were included. Every participant gave his informed consent before participating in the research. On the other hand, those persons diagnosed with serious mental illnesses that may affect their ability to understand or answer the questionnaire were excluded, as well as those with cognitive or communication difficulties that may hinder their participation in the evaluation. Additionally considered as an exclusion criterion is not having performed the test to detect viral load within the period established for this study.

Sample Size Calculation

To determine the sample size, we used a calculator that used a structural equation model approach, in which we anticipated an effect size (factor loading) of 0.5, which would reflect a significant contribution of the latent construct to the observed variables (items), ensuring sufficient construct reliability, a power of 80%, one latent variable, five observable variables, and a probability level of 0.05, resulting in a minimum sample of 100 participants, to detect the specified effect given the structural complexity of this model [15,16].

Variables

The variables considered for this study were age in years, schooling (considered as the last completed grade of studies); municipality of residence (with the category "other" for those participants who indicated not being from a nearby municipality); marital status; employment (was considered positive when they would indicate having an economically remunerated activity); the use of tobacco, alcohol, and illicit drugs (it was considered positive when patients indicated to use in the last 30 days on more than two occasions); the number of pills (including those belonging to a treatment other than ART);

time living with HIV and time on ART (which were calculated considering from the date of HIV diagnosis and the start of ART until the date of enrollment in this study); and finally, clinical stage was classified according to the Center of Disease Control classification [17].

Instrument

The questionnaire consists of 6 questions: the first 4 can be answered with "yes" or "no," while the last 2 require numerical responses: "1. Do you ever forget to take your medication?" "2. Do you always take your medication at the prescribed time?" "3. Do you ever stop taking the drugs if you feel unwell?" "4. Did you forget to take the medication over the weekend?" "5. In the last week, how many times did you miss a dose?" and "6. In the last 3 months, how many full days did you miss taking the medication?" The six questions assess three components of ART adherence: (1) the intentional (question 3), (2) the unintentional (questions 1 and 2), and (3) frequency or quantity (questions 4, 5, and 6). The patient is classified as nonadherent if they answer any in a "nonadherence sense" and if they report missing more than two doses in the last week or reports not having taken more than 2 full days of medication in the last three months [10].

Statistical Analysis

For descriptive analysis, normality tests were performed as necessary. Quantitative variables were described with means or medians, and qualitative variables with frequencies and percentages.

The evidence of validity based on internal structure. was assessed by confirmatory factor analysis using the weighted least squares method with adjusted mean and variance [18]. Model fit was assessed using recognized indices, such as comparative fit index (CFI>0.90) [19], root-mean-square error of approximation (RMSEA<0.08) [20], and weight root-mean-square residual (WRMR<1) [21].

To determine reliability, we estimated the internal consistency coefficients according to Ponterotto and Charter [22]. Scores Cronbach α>0.70 were considered as reliable scores [22]. In addition, we investigated validity by examining the relationship between adherence and undetectable plasmatic viral load using the Pearson correlation coefficient [23].

Finally, Fisher exact tests were used to calculate the odds ratios (OR) to evaluate the association between having an undetectable (patient in control) versus detectable (patient not in control) plasmatic viral load and the scale classification of adherent versus nonadherent (based on the concept that a patient with an undetectable viral load was an adherent patient) [5].

The descriptive and correlational analyses were carried out in SPSS (version 24; IBM Corp) software, while the factor analysis was performed with Mplus (version 7; Muthén & Muthén) software [19]. Finally, the analysis of ORs was performed in Epi Info (Centers for Disease Control and Prevention) software in the Stat Cal module [24].

Ethical Considerations

This research was approved by the Bioethics Committee of the Civil Hospital of Guadalajara (169/23), and it was carried out under the ethical standards for research on human participants. All participants signed an informed consent form that outlined the objectives of the research, the procedures to be followed, and the contact information for the research team. Any questions or concerns were addressed verbally prior to signing the consent form. The information collected was handled exclusively by the research team, and to protect participants' identities an alphanumeric code was assigned to each. Additionally, identifying data were excluded from the database used for analysis. Finally, no direct financial compensation was provided to participants for their involvement in the study.

Results

Characteristics of the Participants

A total of 299 patients were evaluated, 259 men and 40 women, and due to the ratio between groups, we decided to exclude the group of women from the analysis, to maintain the homogeneity and representativeness of the sample [25].

Of the 259 participants included, the mean age was 43 (SD 12) years with an average of 8.97 (SD 6.33) years in ART treatment. Considering schooling, the majority had high school (n=87, 34.6%) or middle school (n=57, 22%) education. The most common municipality of residence was Guadalajara (n=110, 42.5%) and the vast majority were single (n=214, 82.6%). Regarding substance use, 89.6% (n=232) had not consumed illicit substances in the last 30 days while 33.6% (n=87) had consumed alcohol and tobacco in the same period. The mean number of pills per day was 2 (SD 2) and the mean number of years living with HIV was 10 (SD 7.17) with an average CD4+ T lymphocyte count of 686.5 cells/µL (SD 353.88; Table 1).

Table 1. Sociodemographic characteristics of the sample (N=259).

	Values
Age (years), mean (SD)	43 (12)
Schooling, n (%)	
Illiterate	2 (0.77)
Can read and write	16 (6.2)
Primary school	44 (17)
Junior high school	57 (22)
High school	87 (34.6)
University	50 (19.3)
Postgraduate	3 (1.2)
Municipality of residence, n (%)	
Guadalajara	110 (42.5)
Zapopan	32 (12.4)
Tlaquepaque	24 (9.3)
Tonalá	16 (6.2)
Zapotlanejo	2 (0.8)
Tepatitlán	1 (0.4)
La Barca	1 (0.4)
Arandas	2 (0.8)
Zapotlán el Grande	2 (0.8)
Other	69 (26.7)
Marital status, n (%)	
Single	214 (82.6)
Married	16 (6.2)
Common law marriage	22 (8.5)
Widower	2 (0.8)
Regular employment, n (%)	
No	37 (14.3)
Yes	222 (85.7)
Used any illicit substance within the last 30 days, n (%)	
No	232 (89.6)
Yes	27 (10.4)
Consumed alcohol in the last 30 days, n (%)	
No	172 (66.4)
Yes	87 (31.3)
Used tobacco in the last 30 days	
No, n (%)	171 (66)
Yes, n (%)	87 (33.6)
Pills taken per day, mean (SD)	2 (2)
Years in ART[a], mean (SD)	8.97 (6.33)
Years living with HIV, mean (SD)	10.08 (7.17)
CD4+ T lymphocyte count, mean (SD)	686.47 (353.88)

	Values
Clinical stage according to the CDC[b], n (%)	
A1	122 (47.1)
A2	20 (7.7)
A3	4 (1.4)
B1	16 (6.2)
B2	4 (1.5)
B3	4 (1.5)
C1	37 (14.3)
C2	32 (12.4)
C3	20 (7.7)

[a]ART: antiretroviral therapy.

[b]CDC: Centers for Disease Control and Prevention.

Responses From the Participants

Regarding the responses to the questionnaire, 66.6% (n=172) of the respondents reported not forgetting their medication, while 33.4% (n=87) admitted to having forgotten it at least once. On the second question, 80.3% (n=198) stated that they took their medication at the indicated time and 19.7% (n=61) admitted not following this pattern. Regarding whether they stopped taking the drugs if they felt unwell, the great majority (n=242, 92.2%) stated that they did not do so but 7.8% (n=17) admitted having done so at some time.

The majority reported not having forgotten to take their medication over the weekend (n=250, 96.6%). In the last week, the majority (n=242, 93.8%) reported missing only one dose, while 6.2% admitted missing a dose once or twice.

On the other hand, of the responses to question 6, a total of 66.5% (n=174) indicated not having missed taking medication on any full day, while 30.6% (n=174) acknowledged having experienced some degree of medication interruption.

Finally, according to the scores of the questionnaire, 66.6% (n=174) of participants reported being adherent while 33.4% (n=88) reported being nonadherent (Table 2).

Table 2. Distribution of responses to the Simplified Medication Adherence Questionnaire (N=259).

Questions	Participants, n (%)
1. Do you ever forget to take your medication?	
No	172 (66.6)
Yes	87 (33.4)
2. Do you always take your medication at the prescribed time?	
No	61 (19.7)
Yes	198 (80.3)
3. Do you ever stop taking the drugs if you feel unwell?	
No	242 (92.2)
Yes	17 (7.8)
4. Did you forget to take your medication over the weekend?	
No	250 (96.6)
Yes	9 (3.4)
5. In the last week, how many times did you not take a dose?	
1	242 (93.8)
2	16 (5.9)
3	1 (0.3)
6. In the last 3 months, how many full days did you not take your medication?	
0	174 (66.5)
1	51 (20.6)
2	17 (5.9)
3	7 (3.1)
>4	10 (1)
Outcome on adherence	
Adherent	174 (66.6)
Nonadherent	88 (33.4)

Evidence of Validity Based on Internal Structure

The unidimensional model obtained favorable fit indices (CFI=1; RMSEA=0, 90% CI 0-0.085; WRMR=0.072), as well as factor loadings around what was expected, except for item 2 (Do you always take your medication at the prescribed time? Table 3). After performing a second analysis without item 2, the results are adequate both at the level of fit indices (CFI=1; RMSEA=0, 90% CI 0-0.103; WRMR=0.047) and factor loadings (>0.50; Table 3).

Regarding reliability, the magnitudes were acceptable for both scores (Cronbach α=0.702) and construct (ω=0.718).

Finally, the correlation of adherence with viral load was statistically significant (r=0.128; P=.04), but not with recent CD4 T lymphocyte counts (r=0.015; P=.81).

On the relationship between undetectable viral load and adherence, those patients classified as adherent by the scale were 3 times more likely to be undetectable compared to those classified as nonadherent (OR 3.31, 95% CI 1.13-9.64, P=.04; Table 4).

Table 3. Factor loadings of the Simplified Medication Adherence Questionnaire scale.

Questions	First analysis	Second analysis
1. Do you ever forget to take your medication?	0.815	0.854
2. Do you always take your medication at the prescribed time?	–0.358	—[a]
3. Do you ever stop taking drugs if you feel sick or drink alcohol?	0.612	0.526
4. Did you forget to take your medication over the weekend?	0.383	0.374
5. In the last week, how many times did you not take a dose?	0.677	0.701

[a]Not applicable.

Table 4. Relationship between Simplified Medication Adherence Questionnaire adherence classification and viral load.

	Adherent	Nonadherent	OR[a] (95% CI)	*P* value
Undetectable viral load[b]	168	76	3.31 (1.13-9.64)	.04
Detectable viral load[c]	6	9	—[d]	—

[a]OR: odds ratio.

[b]Plasma viral load of ≤40 copies/mL.

[c]Plasma viral load >40 copies/mL.

[d]Not applicable.

Discussion

Principal Findings

Adherence is vital to achieve viral suppression and increase the quality of life of people living with HIV; however, its measurement represents a major challenge [26]. Although direct methods exist to measure adherence (eg, plasma drug concentration), they are costly and impractical for routine implementation in clinical settings with limited resources such as ours [8], thus, indirect measurement methods based on self-reporting may have a good performance and could be used routinely.

This study aimed to evaluate the reliability and validity of the SMAQ in people living with HIV/AIDS attending a public hospital in the western region of Mexico.

As for the factorial structure, 4 of the 5 original items were retained, and the one that was eliminated obtained the lowest factorial loading, and its response was oriented in the opposite direction to the other items, which reinforces the argument that this type of item usually presents methodological problems. This new four-item version obtained adequate fit and reliability indices (CFI=1; RMSEA=0, 90% CI 0-0.103; WRMR=0.047) [27].

The validity of its relationship with other variables was analyzed using an association with viral suppression (or undetectable viral load), finding a threefold greater probability of achieving viral suppression when classified as adherent (OR 3.31, 95% CI 1.13-9.64, *P*=.04).

The use of the SMAQ in the framework of explanatory studies could provide greater insight into the factors associated with adherence, providing evidence for the design of appropriate interventions to improve adherence in this patient population [28]. This, in turn, would strengthen the capacity of health services to enhance and promote quality care, ultimately yielding a positive long-term impact on people living with HIV.

Concerning previous studies, the study by Agala et al [10] also found adequate factor loadings, except for one item, like this work. However, there is no agreement on the eliminated item, which could be attributed to differences in the populations studied because the analysis of our study was only in men and theirs in women. This highlights the existing differences in adherence behaviors between these groups [10].

This situation underscores the need to consider sociodemographic and cultural characteristics when analyzing the evidence and validity of self-report tools such as the SMAQ.

Limitations

Regarding its limitations, it is important to recognize that, despite its simplicity, it may be subject to self-report bias [11]. Similarly, only men were considered due to the proportion that was recruited from both groups and to maintain statistical homogeneity [29]. This limitation does not allow for the opportunity to explore differences that could influence adherence between the two groups.

It is concluded that the SMAQ presents favorable evidence of validity per internal structure and association with viral suppression, as well as acceptable levels of reliability.

In future studies, we recommend that the SMAQ be analyzed at a psychometric level in the group of women to explore possible differences by gender. In addition, this tool will allow studies focused on adherence and its determinants to be carried out with a greater degree of precision.

Conclusions

The use of the SMAQ provides a valuable tool for assessing adherence among the Mexican population living with HIV. This evaluation is critical for enhancing health outcomes and

optimizing therapeutic interventions. Additionally, the SMAQ helps to strengthen public health programs by furnishing reliable data on treatment adherence, which can inform the development of targeted strategies to support patient engagement and medication management.

Acknowledgments

LEDMT contributed to the conception of the idea, the development of the protocol, analysis of results, and conclusions. LAGH worked on the development of the protocol, processing of results, and writing of this paper. JFAV did the writing of the discussion and the final review. PM-A, AVR, and VVRH selected and recruited participants. JAVR carried out the analysis of results and the development of the protocol. MGHG assisted in the general review of the document and the preparation of the final version. SD-L helped with the analysis of the scale, statistical analysis, and writing of the results.

Conflicts of Interest

None declared.

References

1. Hoja informativa—últimas estadísticas sobre el estado de la epidemia de sida. ONUSIDA. 2018. URL: https://www.unaids.org/es/resources/fact-sheet [accessed 2024-11-26]
2. Boletín de atención integral de personas con VIH. CENSIDA. 2024. URL: https://www.gob.mx/censida/articulos/boletin-de-diagnostico-y-tratamiento-antirretroviral-censida?idiom=es [accessed 2024-11-26]
3. Long-Term HIV treatment adherence for key populations. ONUSIDA. URL: https://www.fhi360.org/wp-content/uploads/drupal/documents/epic-long-term-hiv-adherence-guide.pdf [accessed 2024-11-26]
4. Oliveira RDS, Primeira MR, Santos WMD, Paula CCD, Padoin SMDM. Association between social support and adherence to anti-retroviral treatment in people living with HIV. Rev Gaucha Enferm 2020;41:e20190290 [FREE Full text] [doi: 10.1590/1983-1447.2020.20190290] [Medline: 32667420]
5. Tale AH, Tegegne AS, Belay DB. Predictors of viral load and medication adherence among HIV-positive adults under treatment at Felege-Hiwot Comprehensive Specialized Hospital, North-West, Ethiopia. HIV AIDS (Auckl) 2023;15:477-489 [FREE Full text] [doi: 10.2147/HIV.S422980] [Medline: 37593198]
6. Jorge D, Varela A, De PS, De P, Salud L, Nilza D. Sistema de vigilancia epidemiológica de vih informe histórico de VIH 1er trimestre. 2023. URL: https://www.gob.mx/salud [accessed 2024-11-26]
7. Adherence to Longterm Therapies. Geneva, Switzerland: World Health Organization; 2003.
8. López-Romero LA, Romero-Guevara SL, Parra DI, Rojas-Sánchez LZ. Adherencia al tratamiento: concepto y medición. Hacia la promoción de la salud 2016;21(1):117-137. [doi: 10.17151/hpsal.2016.21.1.10]
9. Knobel H, Alonso J, Casado J, Collazos J, González J, Ruiz I, GEEMA Study Group. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. AIDS 2002;16(4):605-613. [doi: 10.1097/00002030-200203080-00012] [Medline: 11873004]
10. Agala CB, Fried BJ, Thomas JC, Reynolds HW, Lich KH, Whetten K, et al. Reliability, validity and measurement invariance of the simplified medication adherence questionnaire (SMAQ) among HIV-positive women in Ethiopia: a quasi-experimental study. BMC Public Health 2020;20(1):567 [FREE Full text] [doi: 10.1186/s12889-020-08585-w] [Medline: 32345253]
11. Ortega SF, Sánchez PJ, Pérez VM, Pereira PP, Muñoz CM, Lorenzo AD. Validación del cuestionario simplificado de adherencia a la medicación (SMAQ) en pacientes con trasplante renal en terapia con tacrolimus. Nefrologia 2011;31(6):690-696 [FREE Full text]
12. Merino-Soto C, Calderón-De La Cruz GA. Validity of peruvian studies on stress and burnout. Rev Peru Med Exp Salud Publica 2018;35(2):353-354 [FREE Full text] [doi: 10.17843/rpmesp.2018.352.3521] [Medline: 30183913]
13. Elangovan N, Sundaravel E. Method of preparing a document for survey instrument validation by experts. MethodsX 2021;8:101326 [FREE Full text] [doi: 10.1016/j.mex.2021.101326] [Medline: 34434840]
14. Leyva-Moral JM, Loayza-Enriquez BK, Palmieri PA, Guevara-Vasquez GM, Elias-Bravo UE, Edwards JE, et al. Adherence to antiretroviral therapy and the associated factors among people living with HIV/AIDS in Northern Peru: a cross-sectional study. AIDS Res Ther 2019;16(1):22 [FREE Full text] [doi: 10.1186/s12981-019-0238-y] [Medline: 31462291]
15. A-priori sample size calculator for structural equation models. Free Statistics Calculator. 2006. URL: https://www.danielsoper.com/statcalc/calculator.aspx?id=89 [accessed 2024-11-26]
16. Dominguez-Lara S. Propuesta de puntos de corte para cargas factoriales: una perspectiva de fiabilidad de constructo. Enfermería Clínica 2018;28(6):401-402. [doi: 10.1016/j.enfcli.2018.06.002]
17. Casanova CJ, Carmen RFM, Gómez GM. Manejo del paciente con VIH en la atención primaria. Actualización en Medicina Familiar. URL: https://www.comunidad.madrid/sites/default/files/doc/sanidad/manejo_paciente_vih_en_ap.pdf [accessed 2024-11-26]
18. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychol Methods 2002;7(1):64-82 [FREE Full text] [doi: 10.1037/1082-989x.7.1.64] [Medline: 11928891]

19. Muthén LK, Muthén BO. Statistical analysis with latent variables user's guide. 1998. URL: https://www.amazon.com/Mplus-Statistical-Analysis-Variables-Version/dp/B0050JZXQW [accessed 2024-11-26]
20. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21(2):230-258. [doi: 10.1177/0049124192021002005]
21. DiStefano C, Liu J, Jiang N, Shi D. Examination of the weighted root mean square residual: evidence for trustworthiness? Struct Equation Model: Multidiscip J 2017;25(3):453-466. [doi: 10.1080/10705511.2017.1390394]
22. Ponterotto JG, Charter RA. Statistical extensions of Ponterotto and Ruckdeschel's (2007) reliability matrix for estimating the adequacy of internal consistency coefficients. Percept Mot Skills 2009;108(3):878-886. [doi: 10.2466/PMS.108.3.878-886] [Medline: 19725323]
23. Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Prof Psychol Res Pr 2009;40(5):532-538. [doi: 10.1037/a0015808]
24. Epi InfoTM User Guide-Chapter 12-StatCalc 12. StatCalc. 2012. URL: https://www.cdc.gov/epiinfo/pdfs/userguide/12_statcalc.pdf [accessed 2024-11-26]
25. Jager J, Putnick DL, Bornstein MH. II. More than just convenient: the scientific merits of homogeneous convenience samples. Monogr Soc Res Child Dev 2017;82(2):13-30 [FREE Full text] [doi: 10.1111/mono.12296] [Medline: 28475254]
26. Nachega JB, Marconi VC, van Zyl GU, Gardner EM, Preiser W, Hong SY, et al. HIV treatment adherence, drug resistance, virologic failure: evolving concepts. Infect Disord Drug Targets 2011;11(2):167-174 [FREE Full text] [doi: 10.2174/187152611795589663] [Medline: 21406048]
27. Suárez-Álvarez J, Pedrosa I, Lozano L, García-Cueto E, Cuesta M, Muñiz J. Using reversed items in likert scales: a questionable practice. Psicothema 2018;30(2):149-158. [doi: 10.7334/psicothema2018.33] [Medline: 29694314]
28. Duthely LM, Sanchez-Covarrubias AP. Digitized HIV/AIDS treatment adherence interventions: a review of recent SMS/Texting mobile health applications and implications for theory and practice. Front Commun (Lausanne) 2020;5:530164 [FREE Full text] [doi: 10.3389/fcomm.2020.530164] [Medline: 33644162]
29. Girum T, Wasie A, Lentiro K, Muktar E, Shumbej T, Difer M, et al. Gender disparity in epidemiological trend of HIV/AIDS infection and treatment in Ethiopia. Arch Public Health 2018;76:51 [FREE Full text] [doi: 10.1186/s13690-018-0299-8] [Medline: 30237881]

Abbreviations

ART: antiretroviral treatment
CFI: comparative fit index
OR: odds ratio
RMSEA: root-mean-square error of approximation
SMAQ: Simplified Medication Adherence Questionnaire
WRMR: weight root-mean-square residual

Edited by T de Azevedo Cardoso; submitted 15.04.24; peer-reviewed by M Nacher, K Hoti; comments to author 11.10.24; revised version received 01.11.24; accepted 14.11.24; published 07.01.25.

Please cite as:
Del Moral Trinidad LE, González Hernández LA, Andrade Villanueva JF, Martínez-Ayala P, Valle Rodríguez A, Ruíz Herrera VV, Vizcaíno Résendiz JA, Herrera Godina MG, Dominguez-Lara S
Simplified Medication Adherence Questionnaire (SMAQ) for People Living With HIV in a National Hospital in Mexico: Instrument Validation Study
Interact J Med Res 2025;14:e59562
URL: https://www.i-jmr.org/2025/1/e59562
doi:10.2196/59562
PMID:39773472

Original Paper

Interstep Variations of Stairways and Associations of High-Contrast Striping and Fall-Related Events: Observational Study

Sara A Harper[1,2], PhD; Chayston Brown[1], BS; Shandon L Poulsen[1], BS; Tyson S Barrett[3], PhD; Christopher J Dakin[1], PhD

[1]Department of Kinesiology and Health Science, Sorenson Legacy Foundation Center for Clinical Excellence, Utah State University, Logan, UT, United States

[2]Kinesiology Department, University of Alabama in Huntsville, Huntsville, AL, United States

[3]Highmark Health Research Institute, Pittsburgh, PA, United States

Corresponding Author:
Sara A Harper, PhD
Kinesiology Department
University of Alabama in Huntsville
301 Sparkman Drive NW
Huntsville, AL, 35899
United States
Phone: 1 256 824 2184
Email: sah0075@uah.edu

Abstract

Background: Interstep variations in step riser height and tread depth within a stairway could negatively impact safe stair negotiation by decreasing step riser height predictability and, consequently, increasing stair users' fall risk. Unfortunately, interstep variations in riser height and depth are common, particularly in older stairways, but its impact may be lessened by highlighting steps' edges using a high-contrast stripe on the top front edge of each step.

Objective: This study aimed to determine (1) if fall-related events are associated with greater interstep riser height and depth variations and (2) if such fall-related events are reduced in the presence of contrast-enhanced step edges compared with a control stairway.

Methods: Stair users were video recorded on 2 public stairways in a university building. One stairway had black vinyl stripes applied to the step's edges and black-and-white vertical stripes on the last and top steps' faces. The stairway with striping was counterbalanced, with the striped stairway than a control, and the control with stripes. Each stair user recorded was coded for whether they experienced a fall-related event. A total of 10,000 samples (observations) of 20 fall-related events were drawn with 0.25 probability from each condition to determine the probability of observing a distribution with the constraints outlined by the hypotheses by a computerized Monte Carlo simulation.

Results: In total, 11,137 individual stair user observations had 20 fall-related events. The flights that had 14 mm in interstep riser height variation and 38 mm in interstep depth variation were associated with 80% (16/20) of the fall-related events observed. Furthermore, 2 fall-related events were observed for low interstep variation with no striping, and 2 fall-related events were observed during low interstep variation with striping. A total of 20 fall-related events were observed, with 4 occurring on flights of stairs with low interstep variation. For stairs with high variability in step dimensions, 13 of 16 (81%) fall-related events occurred on the control stairway (no striping) compared with 3 of 16 (19%) on the high-contrast striping stairway. The distribution of fall-related events we observed between conditions likely did not occur by chance, with a probability of 0.04.

Conclusions: These data support the premise that a vision-based strategy (ie, striping) may counteract fall risk associated with interstep riser height and tread depth variation. Possibly, perception and action elicited through the horizontal-vertical illusion (striping) may have a positive impact on the incidence of fall-related events in the presence of high interstep riser height and depth variation. The findings of this study suggest that contrast enhancement (ie, striping) may be a simple and effective way to reduce the risk of falls associated with interstep variation, highlighting the potential for this approach to make a significant impact on fall prevention efforts.

(Interact J Med Res 2025;14:e60622) doi:10.2196/60622

KEYWORDS

stairs; stairway safety; riser height; tread depth; horizontal-vertical illusion; fall risk; fall prevention; videos; Monte Carlo simulation; public health; vision-based strategy; health promotion; adults; geriatric

Introduction

When approaching a stairway, stair users seemingly anticipate uniformity in the step riser height and tread depth [1,2]. However, this assumption may lead to fewer attentional resources being allocated to estimating these metrics, potentially compromising the safe negotiation of the stairway. Unfortunately, interstep riser height and tread depth variations are common. Often described as dimensional uniformity in building code, stair riser heights and tread depths shall be of uniform size and shape. The tolerance between the largest and smallest riser height, or tread depth, shall not exceed 3/8 inch in any flight of stairs. In our investigation, the range of dimension uniformity is referred to as interstep variation. Interstep variations, as small as 6 mm, between adjacent stair risers or treads can disrupt step cadence and increase the risk of accidents or falls [3]. One strategy to mitigate this risk involves applying a horizontal-vertical illusion and black stripes to the top front edge of each step, which could potentially decrease the frequency of slips, trips, and falls [4-6]. Recent research has demonstrated that adding a high-contrast stripe along the top front edge of each step [4,7-11] can lead to increases in heel clearance above the step. Similarly, adding vertical monochrome striping to the faces of the bottom and top steps can also enhance vertical foot clearance [4,5,12,13].

Although the exact mechanism behind these interventions is not fully understood, it is possible that they increase step height by drawing more attention to each step's edge [7] or by creating a horizontal-vertical illusion, which makes the steps appear taller than they actually are [14,15]. Ultimately, either mechanism may decrease the likelihood of a slip, trip, or fall on stairs with high interstep variation. We, therefore, hypothesized that when comparing 2 flights of stairs with similar interstep variability, the stairs with vertical monochrome striping and tread edge highlighters would record fewer fall-related events than stairs without this intervention. In addition, stairs with greater interstep variation in riser heights and tread depths would generally record more fall-related events than stairs with less interstep riser height and tread depth variability, but this effect would be lessened with the addition of monochrome striping and tread edge highlighters.

To test our hypotheses, we estimated the probability of observing a range of fall-related event distributions that could plausibly occur by chance, given our hypotheses [high or low interstep riser height and tread depth variations and control or striping intervention conditions]. We codified our hypotheses in a Monte Carlo simulation using 4 constraints based on our hypotheses. First, there would be more fall-related events on control flights of stairs (without the contrast intervention) with greater interstep variation than those with lower interstep variation. Second, both control and intervention (with the contrast enhancement) flights of stairs with low interstep variation should observe a comparable frequency of fall-related events (ie, a difference of less than 2 fall-related events between conditions). Third, fewer (less than or equal to half the number of) fall-related events should be observed with high interstep variation intervention stairs compared with the high interstep variation control stairs. Finally, the relative difference in fall-related events between high and low interstep variation stair flights in the control condition should be greater than on the intervention flights.

Methods

Participants

This cross-sectional study took place on 2 public stairways on the Utah State University campus. Video capture occurred on public stairways, and most stair users appeared to be young adults.

As previously discussed, 11,137 individual stair user observations were coded and balanced across the stairway conditions (control and intervention) and stairways (East and West) [16]. There were 7458 (66.97%) feminine observations and 3679 (33.03%) masculine observations, where most observations (n=10,970, 98.5%) were in the age group of 18-40 years. Additional participant details were described previously [6]. Given the observational nature of this study in a public space, no screening was performed in advance. Eligibility criteria were as follows: (1) inclusion—visually appearing 18 years of age and older, captured within local time (eg, 7 AM-5 PM) and (2) exclusion—use of assistive walking devices (eg, crutches and walking boots), individuals that did not transverse the stairs, or involved unusual stairway behavior, as described by Harper et al [6] were documented and removed during the data cleaning phase.

Protocol

High-resolution security cameras (8 megapixels, 4K Ultra HD, 3840×2160 resolution at 7 frames per second, Lorex cameras [Lorex Technology Inc]) were placed in the stairways to record stair users' behavior.

Intervention

High-contrast black vinyl film (Gerber High-Performance Series 220 vinyl film [Gerber Technology]) stripes (5.5 cm wide, 0.063-0.09 mm thick) were applied flush to the top front edge of each stair [4,9,12]. A total of 19 black-and-white vertical vinyl stripes (12 cycles/1 meter) were placed on the very bottom and top steps' faces [4,9,12].

Figure 1 [6] depicts the combined striping intervention. Stairway interstep riser height and depth variations were measured across every step, from the middle part of the step (Figure 2) [6].

Figure 1. Control and interventional stairway conditions. Intervention conditions are further illustrated to show step riser (face of step) and top of step (superior view) vinyl stripes. Step riser height=step height. Stairway width=step tread depth. Vinyl striping depth. Adapted from Harper et al [6], which is published under Creative Commons Attribution 4.0 International License [17].

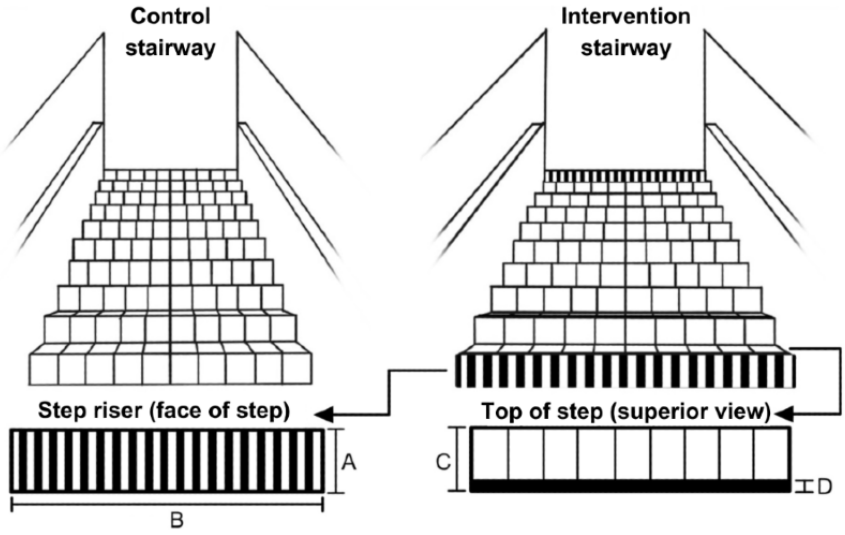
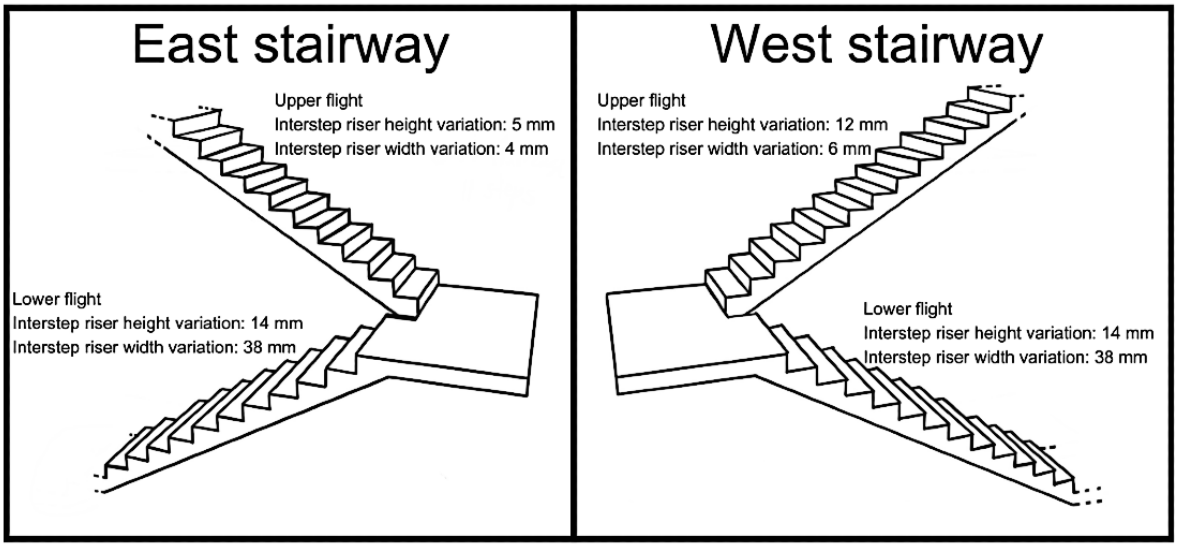


Figure 2. Real-world stairway design. East upper and lower stairway flights interstep riser height variation ranged between 5 mm and 14 mm, and interstep depth between 4 mm and 38 mm, respectively [6]. West upper and lower stairway flights' interstep riser height variation ranged between 12 mm and 14 mm, as well as interstep tread depth variation of 6 mm and 38 mm respectively, are shown [6]. Adapted from Harper et al [6], which is published under Creative Commons Attribution 4.0 International License [17].



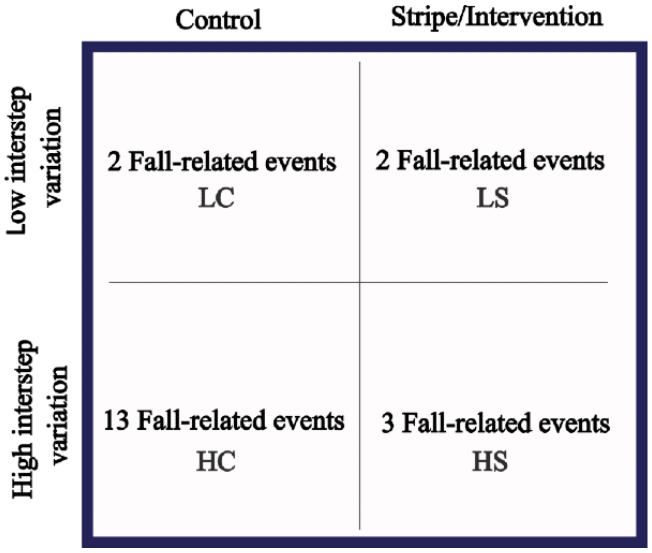
In total, 48 steps were observed across 4 flights of stairs. The control stairway was unaltered and used to compare with the intervention stairway. Halfway through data collection, the intervention (striped) and control stairways were counterbalanced.

Measures

Each outcome variable was assessed by stairway location (East and West) and condition (intervention and control), as well as by stairway flight (lower and upper; Figure 3).

In total, 4 assumptions were used to code our hypotheses based on our a priori knowledge of the total number of fall-related events recorded (n=20): (1) the number of fall-related events high interstep variations, control condition>low interstep variations, control condition; (2) the difference between low interstep variations, stripe intervention and low interstep variations, control condition ≤2 fall-related events; (3) the number of falls in high interstep variations, control condition will be ≥2 times of high interstep variations stripe intervention; and (4) the difference between low interstep variations, control condition and high interstep variations, control condition > the difference between low interstep variations, stripe intervention and high interstep variations stripe intervention. The probability of a distribution meeting these assumptions occurring by chance is approximately 0.04 with a sample size of 20.

Figure 3. The probability of observing a range of distributions of fall-related events. The following assumptions evaluated the observed fall-related event distribution. HC: high interstep variations, control condition; HS: high interstep variations stripe intervention; LC: low interstep variations, control condition; LS: low interstep variations, stripe intervention.



Data Sources

We coded stair users' navigation direction (ascent and descent) and the presence of a fall-related event. As described previously, a fall-related event was coded if an observed stair user experienced a relatively subtle trip or slip, with minimal recovery action, through a complete loss in balance, resulting in a fall [6]. If a fall-related event occurred, the stairway and step number (starting from the bottom to the top) were recorded. To assess research bias or intercoder reliability, each week, one of the researchers would randomly select and evaluate 10% of the data collected for that week. If an error was present, a second coder reviewed all data recorded by the first coder for the week in question, and the second coder made a determination on the final coding record [6]. Data were encoded using Microsoft Excel.

Statistical Analysis

Data are presented as mean (SD) as well as count (percentage) of observed results. The distribution of fall-related events was assessed across stairway flights (interstep variations) and conditions (control and intervention) using a 0.5-inch threshold (approximately 13 mm), given that 75% (80/101) of stair accidents occurred in stairways with interstep riser height variations of ≥0.5 inch [18]. A Monte Carlo simulation in Julia [19] was used to estimate the probability of observing the distribution of fall-related events defined by our hypotheses by chance. Specifically, 10,000 samples (observations) of 20 fall-related events (refer to the Results section) were drawn with 0.25 probability from each condition to determine the probability of observing a distribution with the constraints outlined by the hypotheses through a computerized Monte Carlo simulation [20,21].

Ethical Considerations

Ethical approval was obtained from the institutional review board of Utah State University (10773). As an observational study, participants did not give written consent. Given the observational nature, participants were not compensated to participate. The video recordings were taken in a public setting, and only the approved research team had access to the identifiable data. Therefore, our video data are not available as supplementary material.

Results

Individuals who had any visible health-related conditions (eg, crutches and walking boots) were documented and removed before analysis. In addition, those who visually appeared under the age of 18 years were documented and removed before analysis. All steps were measured at the center of the steps. The average of the East and West stairways combined was 168 (SD 4) mm, with the average riser height of the West stairway steps being 171 (SD 3) mm, and the East stairway being 166 (SD 3) mm, independently. The average step tread depth for the East and West stairways was 328 (SD 8) mm, the average depth of the West stairway steps was 327 (SD 8) mm and the Easy stairway was 329 (SD 8) mm.

Of 20 observed fall-related events, 80% (n=16) of events were observed on the flights where interstep variations were the greatest (riser height ranged 14 mm and tread depth ranged 38 mm). In comparison, 4 of 20 fall-related events were observed on flights with lower interstep variation. Between East and West stairways, 7 of 20 (35%) fall-related events occurred on steps that had step riser heights greater than 1 SD from their mean including 2 fall-related events that occurred on the first step while ascending (step riser height=170 mm; flight mean 166, SD 3 mm), 3 events on the last step while descending the East, lower stairway (height=155 mm; mean 166, SD 3 mm), 1 event on the last step while ascending the East, upper stairway (height=170 mm; mean 166, SD 3 mm), and 1 event on the second-to-last step while ascending the West, lower stairway (height=173 mm; mean 171, SD 4 mm).

A total of 20 fall-related events were recorded. On flights where interstep riser height variation ranged 14 mm and step depth variation ranged 38 mm, 13 of 16 (81%) fall-related events occurred on the control stairway condition (no striping),

compared with 3 of 16 (19%) on the striped intervention stairway. Finally, 2 fall-related events were observed for low interstep variation with no striping, and 2 fall-related events were observed during low interstep variation with striping.

The estimated probability of observing data that fit the range of distributions constrained by our hypotheses and using a sample size of 20 fall-related events, was approximately 0.04 (Figure 3). This result suggests that (1) interstep variations may be contributing to falls and (2) adding a striping intervention to the stairs may reduce the impact of interstep variations on fall-related events.

Discussion

Principal Results

We sought to assess the impact of interstep riser height and tread depth variations on fall-related events (eg, slips, trips, or falls) occurring on stairways to determine if fall risk increases with greater interstep variations but is reduced by adding high-contrast striping. The lower flights of stairs, which had interstep riser height variations that ranged 14 mm and interstep depth variations that ranged 38 mm, accounted for 80% (16/20) of the observed fall-related events, supporting the notion that stairways with greater interstep variations may be associated with a greater risk of fall-related events. Furthermore, 35% (7/20) of fall-related events were observed on steps where interstep riser height variation was greater than 1 SD from the flight mean. Together, these results suggest that flights of stairs with greater interstep riser height variation exhibit more fall-related events than flights with lower interstep variation. When high-contrast striping was added to flights of stairs with high interstep variation, there were fewer fall-related events observed over a similar time period (3/20 fall-related events in the intervention vs 13/20 fall-related events in the control condition). This result suggests the addition of high-contrast striping to stair flights with high interstep variation may reduce the number of fall-related events resulting from interstep variation.

Comparison With Previous Work

Interstep Variations Increase Fall Risk

Interstep variations on stairways can have a profound negative impact on fall risk. Even minor interstep variations, such as a 6 mm variance in riser height [3], can disrupt step cadence and increase the likelihood of a fall, as can interstep variation in tread depth. Furthermore, a review of 80 stairway falls from 1992 to 2007 found that 60% (48/80) of riser height and 34% (27/80) of interstep depth variation were greater than 3/8 inch in a study by Cohen et al [7], and greater (3/8 inch compared with 0.5 inch) interstep variation that can disrupt cadence [3]. Francksen et al [22] found that adults could adjust their stepping behavior for increases in depth, but they could not adjust for interstep riser height variation over 10 mm. Stair-related fall injuries are also more commonly associated with interstep riser height variation than interstep tread depth variation [7]. Alternatively, the greater association of falls with interstep riser height variation could be due to the increased frequency of observing interstep riser heights compared with tread depth variations (eg, > 3/8 inch is observed more often in riser heights [48/80, 60%] than tread depths [27/80, 34%]) [7]. Nevertheless, additional research is needed to explore the extent and impact of interstep variation on stair-related falls. In this study, we observed that 80% (16/20) of fall-related events on stairs occurred when the interstep riser height variation was 14 mm and depth was 38 mm. Since interstep riser height and depth variations were both present in our observational design, we are unable to distinguish which of these factors may have had a greater role.

The Intervention Was Associated With Reduced Fall-Related Events With Interstep Variation

Broadly, our findings indicate that greater interstep variations are associated with an increased frequency of fall-related events. Importantly, the frequency of these fall-related events decreased when the intervention was present. Our data suggest that adding high-contrast stripes may reduce fall-related events on stairs when large interstep variations are present. Of the 20 fall-related events observed, 16 were on the flights of stairs with larger interstep variation, but only 19% (3/16) occurred when the intervention was present, whereas 81% (13/16) occurred when the intervention was absent.

Mechanisms Contributing to the Reduction in Fall-Related Events

The intervention may induce a horizontal-vertical illusion by the intervention. This occurs when horizontal and vertical lines of similar length are presented together (like the letter “T”), which results in an illusory sense that the vertical line is longer than the horizontal one. The intervention used in this study included black vinyl stripes applied to each step’s top front edge and black-and-white vertical stripes on the face of the first and last steps. Together, the abutting edges of the combined striping that formed “T”-like configurations could have contributed to an increased perceived step riser height [14], resulting in greater step heights. Indeed, previous research suggests that under similar experimental conditions, perceived step height is increased and is associated with an increase in the height of the step taken [12,13,15]. In further support for this mechanism, the horizontal-vertical illusion effect is reduced when only edge highlighters are present [13], suggesting that the vertical lines contribute to the increased step height and, perhaps, that the horizontal-vertical illusion is a primary mechanism contributing to the reduction in fall-related events we observed with the intervention. It is also possible that other intervention formats, such as those that could induce the Müller-Lyer illusion [23], could reduce fall-related events. The Müller-Lyer illusion occurs when the perceived length of a line is influenced by the orientation of arrow-like segments attached to its ends. Lines of the same length appear shorter or longer depending on whether the arrowheads at the ends point inward or outward. By manipulating the direction of the fins at the end of the lines, perhaps step height could be increased or decreased depending on the particular interstep variation. Recent outdoor observational research suggests that greater vertical foot clearance occurs when a “fins out” configuration is applied to a 2-step stairway [24], which is the illusion’s expected effect.

Future research could examine whether a "fins in" configuration reduces step height.

By accentuating the steps' edges, the intervention may draw attention to them, thereby enhancing the accuracy with which stair users can estimate the dimensions of each step and then compensate for irregularities. Future research could investigate whether the presence of striping increases awareness of interstep variation, thereby evaluating this potential mechanism. In addition, the novelty of the intervention itself may have drawn attention to the stairs. Schomaker and Meeter [25] suggest that novel visual stimuli, such as a change in contrast, may increase attention to the stairs [26]. It is worth noting that the striping intervention was installed several days before data collection began. It is likely that some stair users had exposure to the striping before the start of video recording and this early exposure could have reduced the novelty of the striping effect before data collection, thus reducing the drawing effect. We also expect that this effect would have worn off over time as many of the observed stair users traverse the stairs frequently due to regular classes in the building.

What Should Be Done About Excessive Interstep Variations

Maybe the most important challenge associated with observing increased fall-related events with greater interstep variations is what to do about it. Is the risk of falls sufficient to warrant widespread evaluation and enforcement of building codes? Although we do not provide recommendations here, conducting assessments, enforcing regulations during construction, and evaluating older stairways (likely to exhibit the greatest interstep variation) may reduce fall-related events, especially when considering cost-effectiveness. For older stairways, as used in this study (built in 1971, and at the time of construction, the 1967 Uniform Building Code [Sec 3305 (d)] that was in place in the United States required the maximum interstep variation in riser height and tread depth to be no more than 3/16 inch), adding an intervention like painting stripes might be the most cost-effective way (an estimated US $288 for the high-contrast striping used here) to reduce the impact of interstep variations (assuming the application or materials used to apply the striping do not themselves increase fall risk through reduced or increased friction, or materials peeling). While increased interstep variation is associated with a greater risk of fall-related events, there are cost-effective interventions that can help reduce this risk. By enforcing building codes and evaluating stairways for interstep variation, we can work toward creating safer environments for everyone.

Limitations

While these results are promising, we acknowledge several limitations of this study. Even with over 10,000 observations, we only observed 20 fall-related events. A larger sample of falls would provide a more precise estimate of the differences between flights of stairs and strengthen the inferences that could be drawn from these data. Since most of the observations in this study were younger adults, future work should consider targeting older or clinical populations (eg, those with visual impairment and mobility-related limitations) to determine if such an intervention could reduce fall-related events. However, future designs will need to consider comparing historical fall frequency records to future intervention fall frequencies rather than using a control condition if it could pose a fall risk to these populations. Furthermore, we did not include Cohen κ, as a measure of interrater reliability. In addition, the lack of a validated questionnaire, such as Yang et al [27], are methodical limitations and should be included in future studies. In addition, given the emphasis on younger adults in this study, it is unknown whether the striping intervention's impact is greater in younger versus older adults. Finally, since the steps used here had interstep variations in both riser heights and tread depths, future observational designs could assess these 2 factors independently to determine the impact of each on fall risk.

Future Directions

Given the considerable negative impact of falls on public health, continued research is necessary to improve safety on stairways. In addition, programs, such as educational campaigns, could be used to raise awareness of factors that contribute to falls [28] and perhaps to help motivate small actions, such as painting stripes on problematic stairways that could have a big impact on public health and provides support for scaling up effective public health interventions for long-term population health benefits.

Conclusions

This study highlights the importance of addressing interstep variations in stairways to reduce the risk of fall-related events. By understanding the factors that contribute to fall risk and implementing cost-effective interventions, we can work toward creating safer environments for everyone. The findings of this study suggest that contrast enhancement (ie, striping) may be a simple and effective way to reduce the risk of falls associated with interstep variations, highlighting the potential for this approach to make a significant impact on fall prevention efforts.

Acknowledgments

We would like to thank Christopher Long, Samantha Corbridge, Alex Braeger, Brevin J. Zollinger, Kenneth Harrison, Travis Boman, Amy Hale, Emmalee Rolfe, McKay Wilding, Erika Larson for their assistance in data coding.

This work was supported by the Utah State University Undergraduate Research and Creative Opportunity Award (2020).

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available since the video data are not deidentified.

Conflicts of Interest

None declared.

References

1. Archea J, Collins BL, Stahl FI. Guidelines for stair safety, Building Science Series, National Institute of Standards and Technology, Gaithersburg, MD; 1979. 1979 May. URL: https://doi.org/10.6028/Nbs.bss.120 [accessed 2024-10-03]
2. Nemire K, Johnson DA, Vidal K. The science behind codes and standards for safe walkways: changes in level, stairways, stair handrails and slip resistance. Appl Ergon 2016 Jan;52:309-316. [doi: 10.1016/j.apergo.2015.07.021] [Medline: 26360223]
3. Templer J. The Staircase. Second Edition. Cambridge: Massachusetts Institute of Technology; 1992.
4. Elliott DB, Foster RJ, Whitaker D, Scally AJ, Buckley JG. Analysis of lower limb movement to determine the effect of manipulating the appearance of stairs to improve safety: A linked series of laboratory-based, repeated measures studies. In: Public Health Research. Southampton (UK): NIHR Journals Library; 2015.
5. Elliott DB, Vale A, Whitaker D, Buckley JG. Does my step look big in this? A visual illusion leads to safer stepping behaviour. PLoS One 2009;4(2):e4577 [FREE Full text] [doi: 10.1371/journal.pone.0004577] [Medline: 19240790]
6. Harper SA, Long C, Corbridge S, Barrett TS, Braeger A, Zollinger BJ, et al. Stairway visual contrast enhancement to reduce fall related events. Journal of Ergonomics 2022;12(303). [doi: 10.35248/2165-7556-22.12.303]
7. Cohen J, LaRue CA, Cohen HH. Stairway falls an ergonomics analysis of 80 cases. Professional Safety 2009;54(1) [FREE Full text]
8. den Brinker BPLM, Burgman LJ, Hogervorst SMJ, Reehorst SE, Kromhout S, van der Windt J. The effect of high-contrast marking of treads on the descent of stairways by low-vision people. International Congress Series 2005 Sep;1282(9):502-506. [doi: 10.1016/j.ics.2005.05.165]
9. Foster RJ, Hotchkiss J, Buckley JG, Elliott DB. Safety on stairs: influence of a tread edge highlighter and its position. Exp Gerontol 2014 Jul;55:152-158 [FREE Full text] [doi: 10.1016/j.exger.2014.04.009] [Medline: 24768822]
10. Thomas NM, Skervin TK, Foster RJ, Parr JV, Carpenter MG, O'Brien TD, et al. Influence of step-surface visual properties on confidence, anxiety, dynamic stability, and gaze behaviour in young and older adults. Hum Mov Sci 2021 Jun;77:102774. [doi: 10.1016/j.humov.2021.102774] [Medline: 33676032]
11. Zietz D, Johannsen L, Hollands M. Stepping characteristics and Centre of Mass control during stair descent: Effects of age, fall risk and visual factors. Gait Posture 2011 Jun;34(2):279-284. [doi: 10.1016/j.gaitpost.2011.05.017] [Medline: 21680187]
12. Foster RJ, Whitaker D, Scally AJ, Buckley JG, Elliott DB. What you see is what you step: the horizontal-vertical illusion increases toe clearance in older adults during stair ascent. Invest Ophthalmol Vis Sci 2015 May;56(5):2950-2957. [doi: 10.1167/iovs.14-16018] [Medline: 26024081]
13. Skervin TK, Thomas NM, Schofield AJ, Hollands MA, Maganaris CN, Foster RJ. The next step in optimising the stair horizontal-vertical illusion: Does a perception-action link exist in older adults? Experimental Gerontology 2021 Jul;149:111309. [doi: 10.1016/j.exger.2021.111309] [Medline: 33716111]
14. Schofield AJ. Second-order texture gratings produce overestimation of height in depictions of rectangles and steps. Vision Res 2022 Nov;200:108101 [FREE Full text] [doi: 10.1016/j.visres.2022.108101] [Medline: 35908371]
15. Skervin TK, Thomas NM, Schofield AJ, Hollands MA, Maganaris CN, O'Brien TD, et al. Using a stair horizontal-vertical illusion to increase foot clearance over an inconsistently taller stair-riser. PLoS One 2021;16(9):e0257159 [FREE Full text] [doi: 10.1371/journal.pone.0257159] [Medline: 34520496]
16. Brown CB, Barrett TS, Long C, Corbridge S, Braeger A, Zollinger BJ, et al. Step edge highlighters and illuminance changes influence stair descent in a real-world setting. Ergonomics 2022 Nov 04;66(9):1219-1228. [doi: 10.1080/00140139.2022.2141347] [Medline: 36314061]
17. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2025-01-03]
18. Miller JA, Esmay ML. Nature and causes of stairway fall. 1961. URL: https://elibrary.asabe.org/abstract.asp?aid=41024&t=2&redir=&redirType= [accessed 2024-12-03]
19. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A Fresh Approach to Numerical Computing. SIAM Rev 2017 Jan;59(1):65-98. [doi: 10.1137/141000671]
20. Kroese DP, Brereton T, Taimre T, Botev ZI. Why the Monte Carlo method is so important today. WIREs Computational Stats 2014 Jun 20;6(6):386-392. [doi: 10.1002/wics.1314]
21. Mundform DJ, Schaffer J, Kim MJ, Shaw D, Thongteeraparp A, Supawan P. Number of Replications Required in Monte Carlo Simulation Studies: A Synthesis of Four Studies. J. Mod. App. Stat. Meth 2011 May 01;10(1):19-28 [FREE Full text] [doi: 10.22237/jmasm/1304222580]
22. Francksen NC, Ackermans TMA, Holzer D, Ebner SA, Maganaris CN, Hollands MA, et al. Negotiating stairs with an inconsistent riser: Implications for stepping safety. Appl Ergon 2020 Sep;87:103131. [doi: 10.1016/j.apergo.2020.103131] [Medline: 32501252]

23. Müller-Lyer FC. Optische urteilstauschungen. Archiv fur anatomie und physiologie. Physiologische Abteilung 1889;2:263-270 [FREE Full text]
24. Shim J, Sims B, Hudnall JT, Rylander J. The effects of Müller-Lyer illusion on toe clearance height in stair ascent. Gait & Posture 2022 Oct;98:78-84. [doi: 10.1016/j.gaitpost.2022.08.019] [Medline: 36063571]
25. Schomaker J, Meeter M. Novelty Enhances Visual Perception. PLoS ONE 2012 Dec 5;7(12):e50599 [FREE Full text] [doi: 10.1371/journal.pone.0050599] [Medline: 23227190]
26. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol 2015 Apr;13(4):e1002128 [FREE Full text] [doi: 10.1371/journal.pbio.1002128] [Medline: 25901488]
27. Yang Y, Feldman F, Leung PM, Scott V, Robinovitch SN. Agreement between video footage and fall incident reports on the circumstances of falls in long-term care. J Am Med Dir Assoc 2015 May 01;16(5):388-394. [doi: 10.1016/j.jamda.2014.12.003] [Medline: 25669670]
28. Harper SA, Corbridge S, Long C, Barrett TS, Braeger A, Zollinger BJ, et al. Safe stairway negotiation: Role of distractions and handrail use. J Safety Res 2022 Sep;82:314-322. [doi: 10.1016/j.jsr.2022.06.007] [Medline: 36031259]

Edited by T de Azevedo Cardoso; submitted 16.05.24; peer-reviewed by S Rietdyk, A Gebre; comments to author 17.09.24; revised version received 28.09.24; accepted 21.11.24; published 08.01.25.

Please cite as:
Harper SA, Brown C, Poulsen SL, Barrett TS, Dakin CJ
Interstep Variations of Stairways and Associations of High-Contrast Striping and Fall-Related Events: Observational Study
Interact J Med Res 2025;14:e60622
URL: https://www.i-jmr.org/2025/1/e60622
doi:10.2196/60622
PMID:

Original Paper

Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study

Hurmat Ali Shah[1], PhD; Mowafa Househ[1], PhD

Hamad Bin Khalifa University, Doha, Qatar

Corresponding Author:
Hurmat Ali Shah, PhD
Hamad Bin Khalifa University
Education City
Doha
Qatar
Phone: 974 70620740 ext 00
Email: hshah@hbku.edu.qa

Abstract

Background: Loneliness is a global public health issue contributing to a variety of mental and physical health issues. It increases the risk of life-threatening conditions and contributes to the burden on the economy in terms of the number of productive days lost. Loneliness is a highly varied concept, which is associated with multiple factors.

Objective: This study aimed to understand loneliness through a comparative analysis of loneliness data on Twitter and Reddit, which are popular social media platforms. These platforms differ in terms of their use, as Twitter allows only short posts, while Reddit allows long posts in a forum setting.

Methods: We collected global data on loneliness in October 2022. Twitter posts containing the words “lonely,” “loneliness,” “alone,” “solitude,” and “isolation” were collected. Reddit posts were extracted in March 2023. Using natural language processing techniques (valence aware dictionary for sentiment reasoning [VADER] tool from the natural language toolkit [NLTK]), the study identified and extracted relevant keywords and phrases related to loneliness from user-generated content on both platforms. The study used both sentiment analysis and the number of occurrences of a topic. Quantitative analysis was performed to determine the number of occurrences of a topic in tweets and posts, and overall meaningful topics were reported under a category.

Results: The extracted data were subjected to comparative analysis to identify common themes and trends related to loneliness across Twitter and Reddit. A total of 100,000 collected tweets and 10,000 unique Reddit posts, including comments, were analyzed. The results of the study revealed the relationships of various social, political, and personal-emotional themes with the expression of loneliness on social media. Both platforms showed similar patterns in terms of themes and categories of discussion in conjunction with loneliness-related content. Both Reddit and Twitter addressed loneliness, but they differed in terms of focus. Reddit discussions were predominantly centered on personal-emotional themes, with a higher occurrence of these topics. Twitter, while still emphasizing personal-emotional themes, included a broader range of categories. Both platforms aligned with psychological linguistic features related to the self-expression of mental health issues. The key difference was in the range of topics, with Twitter having a wider variety of topics and Reddit having more focus on personal-emotional aspects.

Conclusions: Reddit posts provide detailed insights into data about the expression of loneliness, although at the cost of the diversity of themes and categories, which can be inferred from the data. These insights can guide future research using social media data to understand loneliness. The findings provide the basis for further comparative investigation of the expression of loneliness on different social media platforms and online platforms.

(Interact J Med Res 2025;14:e49464) doi:10.2196/49464

KEYWORDS

health informatics; loneliness informatics; loneliness theory; health effects; loneliness interventions; social media; lonely; loneliness; isolation; mental health; natural language processing; tweet; tweets; comparative analysis

Introduction

Loneliness not only affects quality of life but also leads to other mental health issues, thus burdening the public health service system. The monetary loss as a result of loneliness is estimated to be between US $8074.80 and US $12,077.70 per person per year in the United Kingdom [1]. The monetary cost of lost days and loss in productivity is estimated to be US $3.14 billion per year for employees in the United Kingdom. Loneliness is shown to be associated with a high risk for multiple health conditions such as physical and mental health issues, dementia, and early mortality [2].

Loneliness is formally defined as, "the unpleasant experience that occurs when a person's network of social relationships is deficient in some important way, quantitatively or qualitatively" [3]. Loneliness, thus, is the perceived and subjective dissonance between one's desired and actual social contacts and relationships. It is difficult to know whether a person is lonely when there is no direct reporting by the person, as loneliness is a very subjective phenomenon. As opposed to social isolation, which is objective and points to a lack of any social connection, loneliness is hard to identify. Loneliness also follows a U-shaped curve demographically, that is, young and old people are the loneliest [4]. Because social media and technology are more frequently used by the younger generation, they are helpful for analyzing loneliness. A previous report provided a theoretical explication of loneliness, loneliness informatics, and the gaps in research on the topic of loneliness informatics, that is, the application of informatics tools to study loneliness [5].

The causes of mental health issues can vary from genetic factors to social and economic factors, and might involve immediate family or finding meaning in life. The result can be withdrawal from human bonds and touch. Loneliness can be addressed through different interventions. However, there is a need to understand the prevalence of loneliness to devise technology-based and community-oriented strategies. Technology may have resulted in fragmented and individualized existence, but it can also be a great healer. The rise of social media has transformed the way we interact with others, offering new opportunities for social connection and communication. Loneliness is a common experience that can have negative effects on mental and physical health, and social media use has been implicated as a potential contributor to loneliness.

To better understand loneliness, this study performed a comparative analysis of Twitter and Reddit data. The aim of this study was to analyze data from these 2 different social media platforms to identify the topics and themes highly associated with the mention of loneliness. We aimed to identify the associations of socioeconomic, political, and personal-psychological factors with the feeling and expression of loneliness. We were not interested in finding out how many people are lonely. We wanted to assess what kind of expression is associated with loneliness. Both platforms are popular social media sites that allow users to post and interact with others, but they differ in terms of their user base and content focus. Twitter is a micro-blogging site with a broad user base that covers a wide range of topics, while Reddit is a forum-based site with niche communities focused on specific topics.

Loneliness is a subjective experience. Social media has become one of the best sources to study loneliness because loneliness must be studied by self-reporting of this subjective feeling. Understanding the nature of social media platforms viz-a-viz the kind of expression about loneliness is important. This study is a step toward understanding the difference in how loneliness is expressed via different social media platforms. The aim is not to conceptualize loneliness per se as this is beyond the scope of this study, but our analysis will provide insights into what terms, topics, and categories occur more frequently in the mention of loneliness.

The aim of the study is not to discuss the qualitative difference between Reddit and Twitter. The difference between both platforms will stand for all topics of analysis. We do not claim to present the differences between the 2 platforms, which is beyond the scope of this paper. Moreover, discussing the differences between different social media platforms may not be of research interest. The research interest is to assess whether the expression of loneliness varies on different social media platforms. We found that there was some variance. The actual reasons may not be known as we are not aware of any research that mentions why people prefer one social media platform over another. Using natural language processing techniques, we analyzed user-generated content on both platforms to identify and extract keywords and phrases related to loneliness. We then compared the prevalence and nature of loneliness-related content on Twitter and Reddit to identify common themes and trends related to loneliness across the 2 platforms. We wanted to understand the topics mostly associated with loneliness and their frequency of occurrence with the mention of loneliness. Before carrying out this analysis, sentiment analysis of tweets was performed through the valence aware dictionary for sentiment reasoning (VADER) tool from the natural language toolkit (NLTK) [6]. The research questions were as follows:

1. What topics and themes can be identified from the data collected on the expression of loneliness from social media? It is important to understand what the data covers and then to carry out an objective comparison.
2. Are there differences in the expression of loneliness on different social media platforms?
3. How the topics and themes associated with loneliness across social media platforms can be divided into various categories?

Social media data with public access are available for Twitter and Reddit. Facebook, a popular social media platform, does not provide access to information based on key terms. As of October 2023, Facebook provides access to an Ad Targeting dataset. Quora, which is a blog-based social media platform, does not have any public application programming interface (API) to provide access to its information. Reddit and Twitter were selected as their data were publicly available when we started collecting data.

Reddit and Twitter as well as other social media platforms can have their own audience, and it cannot be said with any certainty why people use or prefer one platform over the other. This paper

does not provide an answer to this question. The characteristics of each social media platform are broadly true for any kind of data extracted from them. In this study, we wanted to investigate whether the expression of loneliness varies on different social media platforms. The objective reasons or causal reasons for the differences are beyond the scope of this study.

Several systematic and scoping reviews have been performed to better understand loneliness. Understanding loneliness theoretically and evaluating its relation to mental health have been the topics of several studies [7]. Theoretical studies on loneliness have established the negative effects of loneliness. Similarly, from the health informatics perspective, there have been multiple systematic reviews of the connection between loneliness and mental health [8]. Moreover, there have been studies dealing with the application of technology-based interventions to cope with loneliness [5].

A comparative methodology has been used to compare countries in terms of the parameters of public health [9,10]. Although this study does not involve a country-wise comparative analysis, it is pertinent to note that a comparative analysis of different topics is an established practice. Foufi et al [11] used Reddit as a source to mine textual information about chronic diseases, the entities for the expression of diseases and conditions, and their relations. Similarly, Reddit was used by Schrading et al [12] for an analysis of domestic abuse. They developed a classifier based on the data available on Reddit to classify whether a post is about domestic abuse. Twitter data were analyzed by Tsai and Wang [13] to evaluate people's attitudes toward public health policies. The use of Twitter to analyze mental health or other public health conditions during COVID-19 is a prevalent approach, with studies focusing on a range of issues from insights into the pandemic [14] to analyzing social network data [15] and discerning emotions about COVID-19 [16].

Comparison of social media data for different purposes has been carried out in the literature because each social media platform is associated with a particular kind of user and a particular mode of expression. Cinelli et al [17] used Gab, Facebook, Reddit, and Twitter to assess the difference of opinion across different platforms. Curiskis et al [18] focused specifically on Twitter and Reddit for document clustering and topic modeling, while Gozzi et al [19] performed a comparative analysis of media coverage on Reddit and Wikipedia. Similarly, both Twitter and Reddit were used to monitor the spread of hateful content in the wake of COVID-19 [20].

This study builds on literature reviews involving a comparative analysis of social media and other such comparative techniques to better understand loneliness. This study attempts to assess how loneliness is expressed on Reddit and Twitter. The reason for this analysis is to clarify the roles of these platforms in research on loneliness. As social media can provide an intimate perspective on the experience and expression of the feeling of loneliness, understanding how different social media can contribute to research on loneliness is important. Loneliness is a subjective experience, and analysis of social media data can play a fundamental role in informing strategies to counter loneliness and understand its dynamics. This study attempts to clarify how different social media platforms can be used to understand loneliness. The findings of this study will be helpful in devising a large framework using social media for understanding loneliness.

Methods

Ethical Considerations

This study was exempt from seeking ethical approval as it analyzed publicly available online content. The study did not use the identity of the people from the data but provided an overall picture based on opinions expressed publicly. Data were collected from the public domain, and the data analysis involved aggregate terms, which were not concerned with individual statements. There were no relevant guidelines from the institutional review boards of the authors' institutions regarding the use of publicly available social media data. In the literature, no ethical approval has been taken for using social media data to perform different assessments.

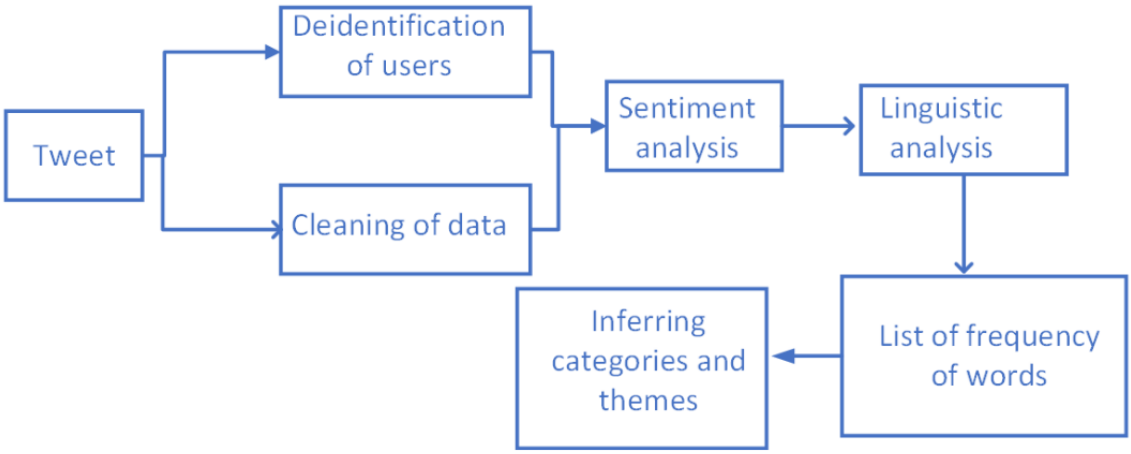
Methodology

We used social intelligence analysis (SIA) to identify the themes associated with the expression of loneliness. SIA is a broad theme that incorporates multiple social media sources such as Facebook, Reddit, and Quora. SIA is important to gain insights into users' data and, in our case, understand the dynamics of loneliness. While SIA can be used for a variety of purposes, such as mining content to create stories and finding trends, we used SIA for sentiment analysis of collected data on loneliness. This is a comparative study to assess the themes associated with loneliness across different social media platforms. Therefore, we used Twitter and Reddit to collect data. The following subsections describe in detail the processes of data collection and analysis for Twitter and Reddit data.

Twitter

We performed an analysis of publicly available data of users posting about loneliness. Figure 1 presents our pipeline of the analysis of data collected from Twitter. Twitter is a social media platform that is used for connectivity and opinion sharing, and allows users to post via short messages consisting of 280 characters. Twitter provides access to user data through its publicly available Twitter API for developers. Relevant tweets about loneliness were gathered and stored in a database. Twitter data analysis involved the following 3 stages: (1) data collection (tweet collection); (2) division of the collected data into negative and other tweets through preliminary analysis (sentiment analysis of tweets); and (3) further analysis of tweets with negative sentiments through manual coding to find relevant themes and categories (manual coding and analysis of tweets).

Figure 1. Pipeline for processing Twitter data.

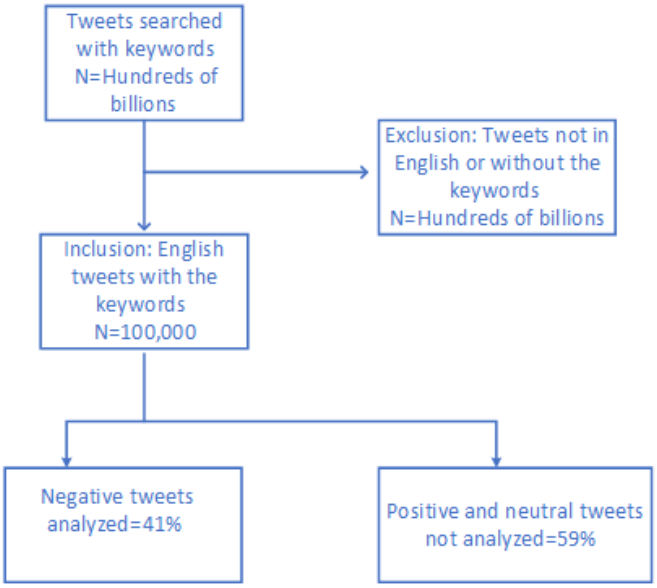


Tweet Collection

Tweets containing the keywords "lonely," "loneliness," "alone," "isolated," and "isolation" were collected. In theoretical literature, the words "loneliness" and "lonely" are used to describe the feeling under consideration in this study. The authors of a previous report [21] collected Twitter data based on the keywords "lonely" and "alone." We went further and included the synonyms and related words of loneliness for collecting our Twitter data.

We did not want to exhaustively search for data from a specific country because we wanted the collected data to be a proof of concept. We focused exhaustively on cities or countries and collected more data. The majority of tweets were from the United States (38%) and India (24%). The rest were from various countries across the globe. Tweets were extracted from these 2 countries to make a subdataset. This was meant to reflect the majority composition of the dataset. The subdataset contained 100,000 tweets. The collected data were merged based on location, user ID, and tweet ID to identify tweets belonging to the United States and India. We did not have control over the countries from which the data were collected. However, after data collection, we found that most tweets were from the United States and India. From the data collected, we made another dataset that contained 60% data from the United States and 40% from India. No city-wise allocation of data or analysis was carried out. Figure 2 presents the process of collecting data from Twitter and the process of tweet analysis.

Figure 2. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram for Twitter data.



Sentiment Analysis of Tweets

The next step was to perform a preliminary analysis of the collected tweets through sentiment analysis. This step is required because loneliness is a negative feeling. To determine how it is expressed in its meaning for mental health analysis, we filtered out tweets where the expression of loneliness was negative. If we decided to report all tweets that contained feelings of loneliness, we would not have required a further step. In our case, the problem becomes determining the association of themes (which may represent loneliness) with the keywords depicting loneliness. For instance, we had to assess the relationship of "hurt," "sick," "tired," "sleep," etc with the expression of loneliness. This task is usually carried out through the association of lexicon categories with tweets including the word "lonely" or "alone." The collected tweets also contained

metaphorical use of lonely or loneliness, which does not pertain to our use of loneliness. Such mentions of loneliness are represented by positive and neutral sentiment tweets. The definition of loneliness in this paper connotates a negative feeling. While loneliness can also be a positive or neutral feeling for some people or some situations, when it comes to its association with mental health issues, the negative consequences of loneliness must be considered.

Therefore, to perform the task of separating tweets with negative connotations for loneliness, we used sentiment analysis. Sentiment analysis was carried out after cleaning the data, such as removing redundant characters, numbers, special characters, user profile IDs, and information like "retweets." For finding tweets with negative sentiment, we used VADER based on Python's NLTK. VADER is suited for microblog content, such as that of Twitter. VADER combines the lexicon, that is, dictionary-based analysis, and rule-based approach to characterize the sentiment. VADER uses gold-standard quality like Linguistic Inquiry and Word Count (LIWC) [22], which has been validated by humans. It distinguishes itself from other efficient tools, such as LIWC, in that it is more sensitive to sentiment expression in social media contexts. VADER also provides valence of the sentiment on a range from 1 to 9. Owing to the sentiment score, we can know through VADER the extent to which the sentiment is negative or positive.

This valence is based on generalizable rules that represent grammatical and syntactical conventions that humans use in contexts meant for emphasizing sentiment intensity. For our purpose, another important feature of VADER is the inclusion of sentiment-bearing lexical nonverbal items, such as emoticons, and verbal items, such as slang, acronyms, and initialisms, which are prevalent in social media contexts. The combination of valence polarity through both the lexicon and rule-based approaches is valuable for fine-grained sentiment analysis. The shortcomings of the lexicon-based approach include coverage, general sentiment intensity, and acquiring a new set of human lexical features.

Manual Coding and Analysis of Tweets

We stored tweets with negative sentiments separately to carry out further analysis. We assessed the themes and categories that most prominently featured in the tweets with negative sentiment. This we performed through manual coding and analysis. We composed a list where the number of occurrences of each word was presented. Before creating the list, we removed stop words and carried out lemmatization to find out the root words. Lemmatization reduces the number of words, and hence, the list was compact.

Manual analysis of the list of occurrences helped in devising larger socioeconomic or emotional-personal categories. The lists were thoroughly searched to identify meaningful words. There were words used for grammatical construction and others not mentioned significantly. This step was subjective and depended on the number of occurrences in the list. The words to include as topics under different categories were guided by the literature [21,23]. The creation of categories and the assignment of topics were subjective. This part of the classification can be considered qualitative as it was guided by the authors' judgments and opinions based on the literature.

Reddit

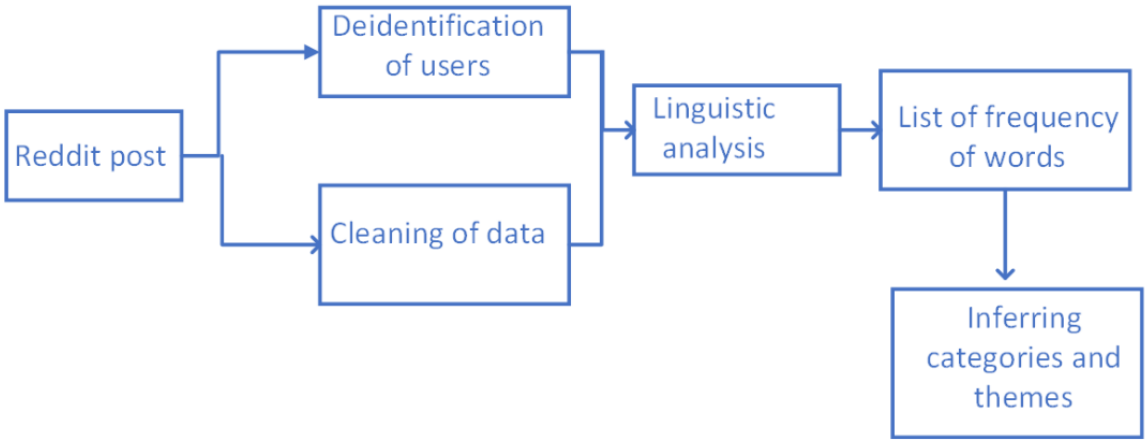
The Reddit assessment involved only 2 steps: data collection and manual coding. The intermediary step was not required because the Reddit data were forum-based, with a focus on coping with the negative effects of loneliness.

Reddit Data Collection

The Reddit data collection methodology was relatively straightforward. Reddit is a forum-based social media platform where people post about a topic on a subforum dedicated to the topic. These subforums are called subreddits. The Reddit API provides access to individual subreddits to download posts along with comments. The API provides access to download the top posts on a topic determined by the number of up-votes and other parameters of engagement. Figure 3 presents our pipeline for processing posts from Reddit. The difference between Figure 1 and Figure 3 is that there is no additional step of sentiment analysis for Reddit posts. After glossing over the subreddit "loneliness," we found that the posts were about the emotional expression of loneliness and did not involve metaphorical or off-topic use of loneliness. The nature of Reddit and subreddits designed for each topic encourages serious engagement on the topics. Because of this fact, sentiment analysis of Reddit data was not deemed important, and posts were analyzed through the frequency of the occurrence of words to identify the themes and topics most highly associated with loneliness.

Reddit posts from the "loneliness" subreddit were collected through the Reddit API. The "loneliness" subreddit had 13,000 members, and people could post and comment in the subforum. Reddit has its own algorithm for providing a score (ie, higher visibility for posts), which also considers inputs from other users in the form of up-voting. We collected the top 2000 Reddit posts from the "loneliness" subforum with all comments. The comments varied for each post in terms of both number and size. It is worth noting that some of the comments were of the same or even larger size than the original post. Thus, comments can be considered to include data on loneliness. The total number of individual texts analyzed was more than 2000 (posts multiplied by the average number of comments). Some posts did not have any comments, and the maximum number of comments for a post in the data we collected was 55. The average number of comments was 4.51, while the total number of comments was 8570. Combining the comments with the posts provided more than 10,000 unique texts or personal expressions on loneliness from Reddit.

Figure 3. Pipeline for processing Reddit data.



Reddit Data Manual Coding and Analysis

We analyzed both the posts and comments to identify the frequency of the occurrence of words to locate the associations of topics and themes with loneliness. The Reddit posts and comments were analyzed without performing any sentiment analysis.

The Reddit posts and comments were collected in a database, and stop words were removed. We performed lemmatization to find out the root words. Then, a list of occurrences of each word was created and analyzed similarly to the method presented in the subsection Manual Coding and Analysis of Tweets.

Reason for Using Manual Coding and Analysis

As the aim of this study was to identify differences between expressions of loneliness on Twitter and Reddit, assessment of important topics was an objective way to achieve this. A qualitative comparison of different tweets and different Reddit posts could be performed, but this would be biased and limited. When the comparison involves thousands of tweets and Reddit posts, it might not be possible to achieve a meaningful subjective comparison. Breaking down the tweets and Reddit posts into topics and categories allows for objective assessments of similarities and differences.

This method of searching for relevant categories of sociopolitical and personal-emotional content and topics was used because it has more flexibility. Usually, the n-gram method indicates the association of words that co-occur, which cannot be perceived to have happened by chance. However, this method does not indicate the occurrence of individual topics and words, and thus, the impact of a topic will not be known. The use of a keyword-based method (ie, using the collected data for some keywords and then reporting their occurrences) is a deductive approach, and it can help in searching the data for predefined keywords. In this study, we used more of an inductive approach, where we let all the important themes and topics in the data emerge. This allowed us to perform a more thorough analysis of data rather than being limited by keywords to look for in the data. This method to analyze the data is a quantitative approach. We did not qualitatively analyze the tweets and Reddit posts (ie, we did not provide our observations on what is contained in them). We reported the number of occurrences of meaningful words and then divided them into relevant categories. This quantitative stating of sentiment analysis and the words contained in negative sentiment tweets under categories was achieved by reading the literature and following the literature.

Results

Comparison of Reddit and Twitter Data

The number of tweets and Reddit posts for analysis was not determined by a specific rule. In the literature, different numbers of tweets and Reddit posts have been used for analysis. For example, Zubaiga [24] carried out an analysis of collected tweets to show the diversity of subjects and dataset sizes. The size of the datasets ranged from over a hundred thousand to more than 10 million. The aim of the study was to reflect in indicative terms the difference in the use of social media according to the expression of loneliness, and therefore, the numbers identified were considered reasonable.

Tables 1 and 2 provide basic details of Twitter and Reddit data, respectively. As discussed earlier, the dataset for Twitter analysis was obtained by combining tweets from the United States and India. The proportions of tweets with negative sentiment for the dataset of each country and for the combined dataset are presented in Table 1. Tweets with negative sentiment were then further analyzed to identify the frequency of the occurrence of topics. For Reddit data, the breakup of the data into words provided a total of more than 25,000 words. For the sake of the meaningful mention of topics and brevity, we only included topics that were mentioned at least 50 times, resulting in 411 words for analysis. Among these 411 words, there were many words related to language construction, such as propositions, and sentence structure. Words that were meaningful in terms of emotions or other expressive qualities were finally included in the analysis.

Table 3 presents the topics highly associated with negative mentions of loneliness in the Twitter dataset. Tweets with negative sentiment were first tokenized and stemmed to obtain a concise list of words and topics associated with loneliness. The list was then analyzed, and meaningful words representing topics of interest, such as emotional, social, and health identifiers, were found. Words like “oh,” “yeah,” and “ur” were ignored when composing the list. For the overall dataset, intimate relationships and interpersonal relationships showed

the highest associations with loneliness (Table 3). “Death” as a topic representing matters of health occurred the most in mentions related to loneliness, along with mentions related to COVID. In terms of socioeconomic factors and political expressions, a wide range of topics were identified.

Table 4 provides a list of themes mentioned with loneliness in the loneliness subreddit. The numbers given represent the total number of occurrences of the word or topic in the Reddit posts. The focus was mostly on relations and emotional expression. As longer posts are allowed, people have more space to express their feelings and open up. Social media platforms provide spaces to own one’s vulnerabilities without facing the backlash that can come in the form of social ostracization. In Table 4, the topics and their relevant thematic categories provide deeper insights into the personal-emotional associations of loneliness.

Table 1. Analysis of Twitter data.

Country	Sentiment analysis of tweets (negative sentiment proportion)
India	33.8%
United States	48.3%

Table 2. Analysis of Reddit data.

Variable	Value, n
Total words	35,057
Words occurring more than 100 times	611
Words occurring more than 1000 times	78

Table 3. Topics highly associated with mentions of loneliness in Twitter data.

Thematic area and topic	Mentions, n
Intimate relationships	
Family	2550
Children	1539
Cheat	64
Woman	3534
Relationship	1213
Interpersonal relationships	
Want	10,452
Need	9854
Feel	5173
Health	
COVID	5715
Die	9652
Life	3729
Socioeconomic factors	
Injustice	1538
Money	1344
Poor	3391
Culture	4452
Colorism (black/white)	2276
Emotional expression/insecurities	
Sad	2343
F*ck	5621
Hate	3542
Political	
War	1328
Protect	1198
Visa	2731
Modi	1576
Citizen	543
Custody	3742
Protest	4068
Insomnia	
Night	7154
Day	4315
Sleep	1564
Awake	543

Table 4. List of themes associated with the expression of loneliness in Reddit data.

Thematic area and topic	Value, n
Intimate relationships	
Love	563
Women	196
Relationship	233
Family	238
Single	111
Friends	1015
Social relations	
Friends	172
Girl	110
She	587
Her	467
People	146
Online	106
Meet	237
Person	428
Interpersonal relations	
Me	2648
He	101
Yours	1596
Us	196
Everyone	245
People	1775
Others	194
Emotional expression	
Thought	197
Hurt	110
Trying	284
Pain	114
Experience	105
Remember	101
Understand	268
Feeling	435
Want	941
Need	539
Feel	1904
Wish	229
Care	262
Self-focused	
I	13,604
Mental	119
My	3989

Thematic area and topic	Value, n
You	5424
Work-related	
Work	331
Job	117
Tried	190
Time	105
School	175
About time	
Life	1608
Year	234
Live	269
Old	145

Similarities and Differences in Reddit and Twitter Data

There are similarities in the categories and themes that can be found in conjunction with mentions of loneliness, but differences in the use of both social media platforms can be crucial in devising strategies to use social media data to understand loneliness. Reddit data are more focused on themes around personal-emotional categories and have a larger number of occurrences of topics around these categories. For Twitter, a range of categories were identified, and personal-emotional categories were the dominant ones, but other topics also occurred. Moreover, the associations of themes and topics with expressions of loneliness were in line with broader psychological linguistic features focused on the self-expression of mental health issues. The difference again was in the range. Twitter can have a wider range of topics that are associated with expressions of loneliness, while Reddit is more focused on psychological personal-emotional categories. These findings will help in providing a guideline for further research into deploying social media data to understand loneliness.

The differences that can be observed in Tables 3 and 4 involve diversity and extensiveness. In Twitter data, the range or diversity of topics and themes can be seen. Because of the limited-character expression on Twitter, people can express their thoughts or opinions, and from these, we can find the association with loneliness. There can be a range of such themes involving direct mentions of loneliness with mentions of a topic in a negative way. On the other hand, Reddit data indicate the extensiveness or depth of a theme or group of topics associated with loneliness.

Some categories, such as “political” and “insomnia,” are missing from Reddit data. The reason for this might be a conjuncture, as direct evidence for the lack of data on Reddit needs to be obtained after following a proper study protocol. However, it can be assumed that because Reddit is more personal (ie, anonymous with the culture being encouraged to reveal personal problems), the general associations of loneliness, such as political and extensive social themes, are not present. For “insomnia,” the easy nature of Twitter (ie, the requirement of short posts) might be seen as a better alternative to discuss sleep problems later at night. Again, these can be conjectural inferences, but the wider social and cultural context can inform the interpretation of these results.

Another reason for Twitter data containing socioeconomic and political categories, which are lacking in Reddit data, is the geographical composition of the data. However, further investigation is needed to determine the geographical composition of the data. According to these findings, Reddit data could be useful for finding the depth of a theme associated with loneliness (ie, what subthemes or topics under a broader category are related to loneliness). Twitter, on the other hand, showed the range and diversity of themes and categories associated with loneliness. This is important to investigate the possible causes of loneliness. The fact that Twitter provides a range of topics and themes and Reddit shows the depth of a theme can be used in a complimentary way.

Discussion

Principal Findings

The study performed a comparative analysis of 100,000 tweets and around 10,000 Reddit posts to understand the themes associated with the expression of loneliness according to socioeconomic and personal-emotional topics. The number of Reddit posts was kept lower than the number of tweets because Reddit posts are lengthier. These 2 social media platforms are important because of their unique nature. Twitter allows short texts, while Reddit allows a freehand composition of one’s opinion and analysis. Analysis of the tweets and Reddit posts provided attributes to understand the associations of socioeconomic and personal-emotional factors with loneliness. These attributes included emotion, sentiment, emojis, and topics. The analysis demonstrated that such attributes could help gather evidence and analyze interactions on the topic of loneliness and other such related topics. The first attribute was emotion, which can serve as a guide in understanding people’s reactions. The second most common attribute was relationships. Other thematic areas, such as health, work, self-focused topics, and

insomnia-related topics, indicated the intimate nature of loneliness. The major finding of this study was that both social media platforms differed in the range of expressions of themes in association with loneliness.

There were similarities in the themes associated with loneliness on social media, but differences in platform usage are crucial for strategies to understand loneliness. Reddit was heavily focused on personal-emotional themes, with a higher frequency of these topics. Twitter, while also emphasizing personal-emotional themes, included a broader range of categories. The thematic associations with loneliness aligned with broader psychological linguistic features centered on mental health self-expression. The key difference was range. Twitter had a wider variety of associated topics, whereas Reddit concentrated on personal-emotional themes. These findings provide guidelines for future research using social media data to understand loneliness.

The methodology developed in this paper showed the association of loneliness with language, which was associated with mental health issues such as anger and depression. The tweets and Reddit posts analyzed prove that psychosocial linguistic features can be found in the self-expression of loneliness, which can identify the dynamics of loneliness [25,26]. Tweets containing keywords associated with loneliness represent a self-focused discourse, which affirms previous literature on loneliness [27,28].

The topics and categories involve linguistic expressions of underlying mental health and related emotions. The findings of this paper regarding the topics and themes associated with loneliness, including mental health, alcoholism, emotional dysregulation, and sociopolitical circumstances, are backed by the literature. The relationships between mental health issues, such as depression and suicidal ideation, and issues related to neurobiology have been discussed by Lam et al [29]. These include conformity with the literature on the association of loneliness with emotional dysregulation and trouble with relationships [30]. The LIWC was used for different thematic concerns for loneliness by previous authors [23]. The study presented in tabular form the related social and emotional supercategories and relevant subcategories of loneliness. Self-referential pronouns were associated with loneliness according to the study. This study also affirms this finding. Our data present such linguistic categories, which have been proven by previous research to be associated with the linguistic nature of loneliness [31].

Finding themes associated with mental health and other issues in public health, such as loneliness, through social media analysis is interesting because such an analysis can provide insights into the first-hand experiences of people. The primary question in this study was “Is there a significant difference in the use of different platforms when it comes to the expression of loneliness?” The answer to this question is not binary. When it comes to larger patterns of the association of topics with loneliness, both platforms can provide similar patterns. However, the difference lies in the range of topics expressed and the depth of a particular topic. This finding is in line with previous findings. A previous study found that Twitter and Reddit can provide different sentiment expressions after a while for opinions regarding vaccination [32].

Which platform is better to use for the analysis of topics associated with loneliness and other related public health issues? From the data analysis and results, we found that Twitter can provide access to a large number of tweets through its API, but special steps in data cleaning and the combination of different data frames have to be taken in order to carry out a meaningful analysis. Reddit, on the other hand, provides access to lengthier posts where users are free from the constraints of space, and multiple insights into the nature of loneliness can be gained from a single post. Therefore, Twitter provides access to a range of insights, while Reddit provides access to a depth of insights. We also found that Reddit posts, although lesser than tweets, provided a lot more topics, but the topics were constrained by the variety of themes and belonged to similar thematic areas.

This study has provided the following insights into the nature and dynamics of loneliness through the collection of data from different social media platforms (Twitter and Reddit):

1. Despite the limitations of the dataset, the varying nature of loneliness can be observed on both platforms. It gives investigators the opportunity to approach the data on the expression of loneliness on different social media platforms differently.
2. Reddit allows users to share their opinions without limits, and Reddit is built around communities. These communities allow people having similar experiences to share their stories in detail. Therefore, Reddit is more suitable for educational and other messaging services, while Twitter can provide a range of topics, which can further be investigated in depth.
3. The data returned by Twitter can be very large, and to use the data, special data cleaning tools and other steps, such as sentiment analysis, should be used. Reddit data are from communities, and as such, they can be directly analyzed. There can be posts from bots and other such automatic or malicious agents. We did not remove these in this study. These need to be removed before further analysis in future work.

Limitations

This study has a number of limitations. First, the data from Twitter and Reddit are self-reported and may not represent the broader population. Only English posts were analyzed, limiting generalizability. Future research should explore various platforms and languages for a more comprehensive understanding of loneliness. Second, the dataset size of both Twitter data and Reddit data was small. We used a limited dataset to determine whether loneliness has different expressions on these platforms. The point was not to be exhaustive but indicative. This limitation can be addressed in a future study by increasing the dataset size to verify or change the findings of this study. Third, the Reddit data were not geo-tagged, and thus, we were not able to determine the countries of the users. However, there are methods to identify the locations of the users of Reddit posts. In the future, these methods can be used to carry out a comparative analysis with geo-located Twitter data. Fourth, there was no control over the data returned by the

Twitter API regarding specifying a country for data collection. The data returned was not probabilistically sampled, and thus, there may be bias in terms of both content and the country or region from which the data are collected. Fifth, the sentiment analysis of Twitter data may have issues. The automatic classification of tweets into negative and positive through sentiment analysis may not be fully accurate. While this was the basis of the study to carry out automated analysis, the results of this automated sentiment analysis need to be validated by looking at a certain number of tweets that were identified as negative. Current models do not detect some linguistic expressions such as sarcasm and humor. New models can be developed, which can correctly classify the sentiment of tweets that use sarcasm and humor. With this approach, we will be able to know the confidence of the analysis and quantify errors. This is related to the methodology of the study, which did not analyze tweets with negative or neutral sentiments. There is a possibility that even a positive or neutral mention of loneliness is associated negatively with mental health, but for the sake of a coherent analysis of a large number of tweets, this has been removed. Validation of sentiment analysis is needed to determine how many tweets with positive sentiment actually had negative connotations.

Conclusion

This study aimed to compare data on loneliness from Twitter and Reddit to identify topics and themes associated with the expression of loneliness. The findings of this study suggest that Twitter provides a wide and diverse range of data related to loneliness, while Reddit shows the depth or extensiveness of a theme associated with loneliness. By analyzing the language used in tweets and Reddit posts, it was possible to identify common themes associated with loneliness, such as social isolation, mental health, and relationship issues. The results of this study provide valuable insights into the ways in which people express and experience loneliness on social media platforms. The ability to identify these themes and topics associated with loneliness could be useful for health professionals and researchers to understand the complex nature of loneliness and its impact on individuals.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

HS designed the study, collected the data, carried out data analysis, and wrote the paper. MH designed and supervised the study.

Conflicts of Interest

None declared.

References

1. Shah S, Nogueras D, van Woerden H, Kiparoglou V. Evaluation of the effectiveness of digital technology interventions to reduce loneliness in older adults: Systematic review and meta-analysis. J Med Internet Res 2021 Jun 04;23(6):e24712 [FREE Full text] [doi: 10.2196/24712] [Medline: 34085942]
2. Hards E, Loades ME, Higson-Sweeney N, Shafran R, Serafimova T, Brigden A, et al. Loneliness and mental health in children and adolescents with pre-existing mental health problems: A rapid systematic review. Br J Clin Psychol 2022 Jun 16;61(2):313-334. [doi: 10.1111/bjc.12331] [Medline: 34529837]
3. Peplau L, Perlman D. Loneliness: A Sourcebook of Current Theory, Research and Therapy. New York, NY: Wiley; 1982.
4. Achterbergh L, Pitman A, Birken M, Pearce E, Sno H, Johnson S. The experience of loneliness among young people with depression: a qualitative meta-synthesis of the literature. BMC Psychiatry 2020 Aug 24;20(1):415 [FREE Full text] [doi: 10.1186/s12888-020-02818-3] [Medline: 32831064]
5. Shah HA, Househ M. Understanding loneliness in younger people: Review of the opportunities and challenges for loneliness interventions. Interact J Med Res 2023 Nov 02;12:e45197 [FREE Full text] [doi: 10.2196/45197] [Medline: 37917125]
6. Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, Michigan, USA p. 216-225. [doi: 10.1609/icwsm.v8i1.14550]
7. Wiwatkunupakarn N, Pateekhum C, Aramrat C, Jirapornchaoren W, Pinyopornpanish K, Angkurawaranon C. Social networking site usage: A systematic review of its relationship with social isolation, loneliness, and depression among older adults. Aging Ment Health 2022 Jul 24;26(7):1318-1326. [doi: 10.1080/13607863.2021.1966745] [Medline: 34427132]
8. Choi HK, Lee SH. Trends and effectiveness of ICT interventions for the elderly to reduce loneliness: A systematic review. Healthcare (Basel) 2021 Mar 07;9(3):293 [FREE Full text] [doi: 10.3390/healthcare9030293] [Medline: 33800099]
9. Khurana A, Rastogi R, Gamperl HJ. A new era of drug safety - New EU pharmacovigilance (PV) legislation and comparison of PV in EU, US and India. International Journal of Pharmacy and Pharmaceutical Sciences 2014;6(7):15-21 [FREE Full text]
10. Sathyanarayanan V, Flowers CR, Iyer SP. Comparison of access to novel drugs for lymphoma and chronic lymphocytic leukemia between India and the United States. JCO Global Oncology 2020 Nov(6):1124-1133. [doi: 10.1200/go.20.00012]

11. Foufi V, Timakum T, Gaudet-Blavignac C, Lovis C, Song M. Mining of textual health information from Reddit: Analysis of chronic diseases with extracted entities and their relations. J Med Internet Res 2019 Jun 13;21(6):e12876 [FREE Full text] [doi: 10.2196/12876] [Medline: 31199327]
12. Schrading N, Alm C, Ptucha R, Homan C. An Analysis of Domestic Abuse Discourse on Reddit. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015 Presented at: 2015 Conference on Empirical Methods in Natural Language Processing; September 17-21, 2015; Lisbon, Portugal p. 2577-2583. [doi: 10.18653/v1/d15-1309]
13. Tsai M, Wang Y. Analyzing twitter data to evaluate people's attitudes towards public health policies and events in the era of COVID-19. Int J Environ Res Public Health 2021 Jun 10;18(12):6272 [FREE Full text] [doi: 10.3390/ijerph18126272] [Medline: 34200576]
14. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: Longitudinal analysis of twitter data. J Med Internet Res 2020 Dec 14;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
15. Ahmed W, Vidal-Alaball J, Vilaseca Llobet JM. Analyzing discussions around rural health on twitter during the COVID-19 pandemic: Social network analysis of twitter data. JMIR Infodemiology 2023 Mar 8;3:e39209 [FREE Full text] [doi: 10.2196/39209] [Medline: 36936067]
16. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
17. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
18. Curiskis SA, Drake B, Osborn TR, Kennedy PJ. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Information Processing & Management 2020 Mar;57(2):102034. [doi: 10.1016/j.ipm.2019.04.002]
19. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: Mixed-methods analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
20. Babvey P, Capela F, Cappa C, Lipizzi C, Petrowski N, Ramirez-Marquez J. Using social media data for assessing children's exposure to violence during the COVID-19 pandemic. Child Abuse Negl 2021 Jun;116(Pt 2):104747 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104747] [Medline: 33358281]
21. Guntuku SC, Schneider R, Pelullo A, Young J, Wong V, Ungar L, et al. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355 [FREE Full text] [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
22. Pennebaker J, Francis M, Booth R. Linguistic Inquiry and Word Count. Mahwah, NJ: Erlbaum Publishers; 2001.
23. Liu T, Ungar LH, Curtis B, Sherman G, Yadeta K, Tay L, et al. Head versus heart: social media reveals differential language of loneliness from depression. Npj Ment Health Res 2022 Oct 18;1(1):16 [FREE Full text] [doi: 10.1038/s44184-022-00014-7] [Medline: 38609477]
24. Zubiaga A. A longitudinal assessment of the persistence of twitter datasets. Asso for Info Science & Tech 2018 May 14;69(8):974-984. [doi: 10.1002/asi.24026]
25. Stravynski A, Boyer R. Loneliness in relation to suicide ideation and parasuicide: a population-wide study. Suicide Life Threat Behav 2001 Jan 07;31(1):32-40. [doi: 10.1521/suli.31.1.32.21312] [Medline: 11326767]
26. Blai B. Health consequences of loneliness: a review of the literature. J Am Coll Health 1989 Jan 09;37(4):162-167. [doi: 10.1080/07448481.1989.9938410] [Medline: 2647810]
27. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct 22;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
28. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. Journal of Research in Personality 2017 Jun;68:63-68. [doi: 10.1016/j.jrp.2017.02.005]
29. Lam JA, Murray ER, Yu KE, Ramsey M, Nguyen TT, Mishra J, et al. Neurobiology of loneliness: a systematic review. Neuropsychopharmacology 2021 Oct 06;46(11):1873-1887 [FREE Full text] [doi: 10.1038/s41386-021-01058-7] [Medline: 34230607]
30. Al-Saggaf Y, Nielsen S. Self-disclosure on Facebook among female users and its relationship to feelings of loneliness. Computers in Human Behavior 2014 Jul;36:460-468. [doi: 10.1016/j.chb.2014.04.014]
31. Buecker S, Ebert T, Götz FM, Entringer TM, Luhmann M. In a lonely place: Investigating regional differences in loneliness. Social Psychological and Personality Science 2020 Apr 10;12(2):147-155. [doi: 10.1177/1948550620912881]
32. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: Comparative study. J Med Internet Res 2022 Oct 17;24(10):e40408 [FREE Full text] [doi: 10.2196/40408] [Medline: 36174192]

Abbreviations

API: application programming interface
LIWC: Linguistic Inquiry and Word Count
NLTK: natural language toolkit
SIA: social intelligence analysis
VADER: valence aware dictionary for sentiment reasoning

Edited by T de Azevedo Cardoso; submitted 30.05.23; peer-reviewed by C Entwistle, T Cahill; comments to author 28.08.23; revised version received 29.10.23; accepted 25.07.24; published 14.03.25.

Please cite as:
Shah HA, Househ M
Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study
Interact J Med Res 2025;14:e49464
URL: https://www.i-jmr.org/2025/1/e49464
doi:10.2196/49464
PMID:

Original Paper

Student and Physician Views of How the Dobbs Decision Affects Training and Practice Location Preferences: Cross-Sectional Questionnaire Study

Morgan S Levy[1*], MPH, MD; Simone A Bernstein[2*], MHPE, MD; Sarah M McNeilly[3], BA; Abigail Liberty[4], MSPH, MD; Shira Fishbach[5], MD; Shikha Jain[6], MD; Jessica A Gold[2], MS, MD; Vineet M Arora[7], MAPP, MD

[1]Department of Radiation Oncology, University of Kentucky College of Medicine, Lexington, KY, United States

[2]Department of Psychiatry, Washington University School of Medicine in St. Louis, St. Louis, MO, United States

[3]Albert Einstein College of Medicine, New York, NY, United States

[4]Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, United States

[5]Department of Obstetrics and Gynecology, University of Michigan School of Medicine, Ann Arbor, MI, United States

[6]Department of Medicine, University of Illinois at Chicago, Chicago, IL, United States

[7]Department of Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Morgan S Levy, MPH, MD
Department of Radiation Oncology
University of Kentucky College of Medicine
800 Rose Street
Lexington, KY, 40536
United States
Phone: 1 9733092423
Email: morgan.levy4@gmail.com

Abstract

Background: By allowing for abortion bans and restrictions to take effect in the majority of US states, the 2022 *Dobbs v Jackson Women's Health Organization* decision portends to have lasting impacts on patient care and the physician workforce. Notably, it is already beginning to impact practice location preferences of US health care workers, evidenced by declining application rates to residency programs in abortion-restrictive states since 2022. Yet, there remains a gap in the literature regarding why this trend exists.

Objective: This study aims to describe what factors are driving the practice location preferences of medical students and physicians after the *Dobbs* decision.

Methods: This study analyzes qualitative data from a web-based, cross-sectional study. In August 2022, a nonprobabilistic sample of physicians and medical students were surveyed on social media about the impact of overturning *Roe v Wade* on practice location preferences, which included the free-text question "Please share your thoughts about the overturning of *Roe v Wade* and how it will affect your decision about your (residency/job or fellowship) programs." A total of 3 independent team members completed an inductive thematic analysis of 524 free responses, resolving differences by discussion.

Results: Approximately 1 in 4 survey respondents also completed the free-response item (524/2063, 25.4%); a total of 219 were medical students, 129 were residents and fellows, and 176 were practicing physicians. Of them, approximately half (261/524, 50.5%) resided in states where abortion bans were in place or anticipated. Those who answered the free-response item were relatively more likely to hail from states with restrictive abortion bans (P<.001) compared to those who did not, with other demographic characteristics being largely similar between the groups. Inductive thematic analysis yielded 2 broad thematic categories: patient-related and workforce-related factors influencing practice decision preferences. The 3 most common themes overall were respondent concerns regarding their patient's access to care (249/524, 47.5%), their desire not to practice or train in a state with abortion restrictions regardless of current residence (249/524, 47.5%), and their personal belief that abortion bans are human rights and/or body autonomy violation (197/524, 37.6%). Some respondents stated that the *Dobbs* decision would not impact their choice of practice location (41/524, 7.8%), and some supported it (35/594, 6.7%).

Conclusions: This study shows that abortion restrictions are having an impact on the practice location preferences of the physician workforce due to both patient care and personal factors. It is important that state policy makers and others who are considering abortion restrictions also consider how to address these concerns of physicians and medical students, to avoid worsening geographic maldistribution of physicians and worsening access to care from physicians for their citizens.

(Interact J Med Res 2025;14:e55035) doi:10.2196/55035

KEYWORDS

abortion; physician workforce; social media; reproductive health; medical education; abortion access; education; survey study; students; training; patient care; care; medical students; human rights; autonomy

Introduction

The US landmark *Dobbs v Jackson Women's Health Organization* decision allowed for widespread restrictions on abortion care, with 14 US states now enforcing total abortion bans and 27 more with bans based on gestational age [1,2]. These include Targeted Regulation of Abortion Providers (TRAP) laws that hamper and criminalize the practice of abortion [2].

While evidence affirms that abortion restrictions have deleterious effects on patient care and public health [3-6], it is important to understand that such policies also impact the health of physicians. A majority of physicians and medical students plan to build families during or after medical training, with thousands desiring pregnancy each year [7,8]. Many rely on infertility treatments, which abortion restrictions hamper [9]. Abortion restrictions, therefore, may deny a significant proportion of the physician workforce comprehensive family planning services, placing them at risk of forced birth [10]. Furthermore, they may also create moral injury among physicians from conflict between personal and professional morals, uncertainty regarding allowable practices, and fear of prosecution [11]. Those who provide abortion care may face increased stigma or even criminalization, depending on the state in which they train or practice. Those who are in restricted states and are not able to provide abortion care may struggle to navigate what is right for their patients versus what is legal, potentially worsening burnout and compassion fatigue [11,12].

Recent analysis from the American Association of Medical Colleges (AAMC) shows that fewer US MD seniors applied for residency positions in abortion-banned states versus nonban states in 2023 [13]. This includes a small but significant decline in the number of applications to obstetrics and gynecology residency programs in restrictive states in 2022 and 2023 [14]. To date, no study has described why physicians hold such preferences. Using an inductive analysis of free-response survey questions from our previous survey, this study aims to describe how state abortion restrictions may influence physicians' and students' decisions about where to live and practice.

Methods

Overview

We conducted a web-based, cross-sectional study for 2 weeks in August 2022. A nonprobabilistic sample of physicians (practicing physicians, fellows, and residents) and medical students were recruited from dedicated physician communities on social media (Twitter [rebranded as X in July 2023], Facebook, and Instagram [Meta Platforms]) through platforms like the American Medical Women's Association and Inside The Match. All physicians and medical students in the United States were eligible to participate, including both those who practice or intend to practice in reproductive health care and those who do not. There was no minimum age for participation. Physicians completed a questionnaire about the impact of overturning *Roe v Wade* on practice location preferences [15]. Respondents reported demographic information and their location preferences for residency (medical students) or fellowship and jobs (physicians). No identifying information was collected.

This analysis focused on the study respondents' stated practice location preferences. Quantitative data from this study were previously published [15]. Survey respondents were offered a free-response question, "Please share your thoughts about the overturning of *Roe v Wade* and how it will affect your decision about your (residency/job or fellowship) programs." An inductive thematic analysis was used [10]. We consulted the Standards for Reporting Qualitative Research to report the study findings [16]. The free-response item was included to allow respondents to contextualize their practice location preferences [17]. The study team is comprised of a medical student pursuing obstetrics and gynecology (OBGYN; SMM), residents in radiation oncology (MSL) and OBGYN (SF), a fellow in Complex Family Planning (AL), and practicing physicians in psychiatry (SAB and JAG) and internal medicine (SJ and VMA). Some team members practice or are training in locations with abortion restrictions, and some practice in less restrictive locations. A total of 4 independent coders (MSL, SAB, SMM, and SF) coded responses until thematic saturation was reached (n=73 for medical students and n=102 for residents, fellows, and practicing physicians) and established the code book through consensus discussion. After establishing the code book, 2 authors coded all responses (n=524), and differences were resolved by discussion. Statistics were done in IBM SPSS (version 29), and group comparisons were calculated with chi-square testing. The CHERRIES checklist for the reporting of internet surveys guided the reporting of the study (Multimedia Appendix 1) [18].

Ethical Considerations

The study was approved as exempt from review by the Institutional Review Board at the University of Chicago (IRB22-1066). Participants provided consent with the opportunity to opt out of the study and were not compensated

for participation. Data were collected without identifiers and are only accessible to study team members.

Results

Demographics

Of the 2063 survey respondents, 524 (25.4%) completed the free-response item. Respondents consisted of medical students (n=219), residents and fellows (n=129), and practicing physicians (n=176). Most identified as cisgender women (391/524, 74.6%). The majority (453/524, 86.5%) of respondents were of reproductive age (less than age 44) and had no children (361/524, 68.9%). Approximately half (261/524, 50.5%) resided in states where abortion bans were in place or anticipated; half (256/524, 49.5%) resided in states where abortion remains legal [19]. Roughly a fifth (114/524, 21.8%) specialized in OBGYN, 13.2% (69/524) specialized in family medicine, and 65.1% (341/524) specialized in another field. The complete demographics of the sample who answered the free-response portion appear in Table 1.

Respondents who answered the free-response item were similar to those who did not by gender (P=.07), race (P=.13), or whether they intended to provide abortion care (P=.22). Respondents in states with restrictive abortion bans (50.5%) were more likely to respond (P<.001) compared with those in a state without restrictive abortion bans (41.7%).

Free-response rates suggest that these qualitative data appropriately represent the spectrum of views on abortion rights and access. Of the overall sample, 82.3% (1698/2063) indicated they would prefer to apply where abortion access is preserved; among them, 23.1% (393) answered the free-response item versus 76.9% (1305/2063) who did not (P<.001). However, of the 9.7% (200/2063) who did not prefer to apply where abortion access was preserved, 41.5% (83) provided a free response, while 58.5% (117) did not (P<.001). Of the 11.1% (229) who indicated that abortion restrictions do not impact their preferences, 32.8% (75) responded versus 67.2% (154) who did not (P<.001).

Table 1. Demographics of medical students (n=219), residents and fellows (n=129), and practicing physicians (n=176) who answered the free response item.

Characteristic	Total (n=524), n (%)	Medical students (n=219), n (%)[a]	Residents and fellows (n=129), n (%)	Practicing physicians (n=176), n (%)
Gender[b]				
Woman	391 (74.6)	158 (72.1)	94 (72.9)	139 (79)
Man	109 (20.8)	48 (21.9)	30 (23.3)	31 (17.6)
Transgender and/or gender nonconforming	7 (1.4)	4 (1.9)	2 (1.6)	1 (0.6)
Prefer to describe	12 (2.3)	2 (0.9)	0 (0)	3 (1.7)
Prefer not to answer	43 (2.1)	7 (3.2)	3 (2.3)	2 (1.1)
Ethnicity[c]				
Hispanic	45 (8.6)	27 (12.3)	10 (7.8)	8 (4.5)
Not Hispanic	456 (87)	181 (82.6)	114 (88.4)	161 (91.5)
Prefer not to answer	23 (4.4)	11 (5)	5 (3.9)	7 (4)
Race[c]				
American Indian, Alaska Native, Native Hawaiian, or other Pacific Islander	2 (0.4)	2 (1)	0 (0)	0 (0)
Asian	49 (9.4)	18 (8.2)	9 (7)	22 (12.5)
Black, African American, or African	37 (7.1)	19 (8.7)	16 (12.4)	2 (1.1)
Multiracial[d]	27 (5.2)	13 (5.9)	6 (4.7)	8 (4.5)
White	361 (68.9)	140 (63.9)	91 (70.5)	130 (73.9)
Prefer to describe	15 (2.9)	8 (3.7)	1 (0.8)	6 (3.4)
Prefer not to answer	33 (6.3)	19 (8.7)	6 (4.7)	8 (4.5)
Sexual orientation				
Bisexual	51 (9.7)	30 (13.7)	8 (6.2)	13 (7.4)
Gay or lesbian	19 (3.6)	8 (3.7)	4 (3.1)	7 (4)
Heterosexual	404 (77.1)	157 (71.7)	106 (82.2)	141 (80.1)
Queer, pansexual, and/or questioning	21 (4)	7 (3.2)	6 (4.7)	8 (4.5)
Don’t know	3 (0.6)	3 (1.4)	0 (0)	0 (0)
Prefer to describe	6 (1.1)	4 (1.8)	0 (0)	2 (1.1)
Prefer not to answer	20 (3.8)	10 (4.6)	5 (3.9)	5 (2.8)
Age range[e] **(years)**				
≤44	453 (86.5)	218 (99.5)	127 (98.4)	68 (38.6)
≥45	71 (13.5)	1 (0.5)	2 (1.6)	108 (61.4)
Relationship status				
Single	128 (24.4)	72 (32.9)	29 (22.5)	27 (15.3)
Partnered	125 (23.9)	89 (40.6)	29 (22.5)	7 (4)
Married	251 (47.9)	53 (24.2)	68 (52.7)	130 (73.9)
Widowed	2 (0.4)	0 (0)	0 (0)	2 (1.1)
Divorced	5 (1)	1 (0.5)	0 (0)	4 (2.3)
Other	5 (1)	1 (0.5)	1 (0.8)	3 (1.7)
Prefer not to answer	8 (1.5)	3 (1.4)	2 (1.6)	3 (1.7)
Children				
Yes	163 (31.1)	25 (11.4)	22 (17.1)	116 (65.9)

Characteristic	Total (n=524), n (%)	Medical students (n=219), n (%)[a]	Residents and fellows (n=129), n (%)	Practicing physicians (n=176), n (%)
No	361 (68.9)	194 (88.6)	107 (82.9)	60 (34.1)
Respondent's current state of residence, by anticipated abortion restriction[f]				
Ban or likely ban[g]	261 (50.5)	125 (58.4)	66 (51.2)	70 (40.2)
Legal[h]	256 (49.5)	89 (41.6)	63 (48.8)	104 (59.8)
Specialties				
Obstetrics and gynecology	114 (21.8)	51 (23.3)	28 (21.7)	35 (19.9)
Family medicine	69 (13.2)	30 (13.7)	17 (13.2)	22 (12.5)
All others	341 (65.1)	138 (63.0)	84 (65.1)	119 (67.6)

[a]Includes US medical students (n=188) and international medical graduates applying to US residency programs (n=31).

[b]Nationally, medical students are 47.9% female and 52.9% male, residents and fellows are 46.8% female and 53% male, and practicing physicians are 35.9% female and 64.1% male [20].

[c]Nationally, medical students are 0.2% American Indian or Alaska Native, 54.6% White, 21.6% Asian, 6.2% Black or African American, 5.3% Hispanic, 8% multiple races, and 3.5% other. Nationally, residents and fellows are 0.11% American Indian or Alaska Native, 48.9% White, 26.6% Asian, 6% Black or African American, 9.2% Hispanic, 4% multiple races, and 3.1% other. Nationally, practicing physicians are 0.1% American Indian or Alaska Native, 63.9% White, 19.2% Asian, 3.6% Black or African American, 5.5% Hispanic, 2% multiple races, and 5.6% other [20,21].

[d]Respondents who selected more than one option are considered multiracial for the purpose of this study.

[e]Age 15-44 years is defined as reproductive age per the Centers for Disease Control and Prevention [22].

[f]Includes all 50 states, Puerto Rico, and the District of Columbia. Excludes the 7 respondents who indicated "other" on their location [19].

[g]Alabama, Arizona, Arkansas, Florida, Georgia, Idaho, Indiana, Iowa, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Montana, North Carolina, North Dakota, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, West Virginia, Wisconsin, and Wyoming [19].

[h]Alaska, California, Colorado, Connecticut, Delaware, District of Columbia, Hawaii, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Vermont, Virginia, and Washington [19].

Overview of the Inductive Analysis of Free-Response Survey Answers

There were 2 groups of themes and 2 stand-alone themes. One group described how practice location decisions impact patient care (Table 2), and the other captured workforce-related concerns (Table 3). The remaining themes included no impact and antiabortion sentiment.

Table 2. Patient factors influencing decisions about practice location emerging from the inductive analysis of the following: for students applying to residency, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your residency application and ranking decisions below," and for fellows and practicing physicians, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your decision about your job or fellowship programs" among respondents (n=524).

Theme	Students (n=219), n (%)	Example quote	Physicians (n=305), n (%)	Example quote
Patient access to abortion (or reproductive care)	84 (38)	I'm horrified when I imagine taking care of a teenager who is being forced to carry out a pregnancy. I'm terrified of the burden of caring for a NICU filled with babies who were born despite having anomalies that make their short lives painful. I can only hope I'm not assaulted or become pregnant without the option to terminate.	165 (54)	I want to be able to support my patients to make good decisions about pregnancy. I need to be able to refer people if they need termination of pregnancy. It goes against my ethics to have to deprive someone of that option. I care foremost about my patients. If one of my patients died because she couldn't get an abortion, I wouldn't be able to live with myself.
Did not want politics to interfere with medical care decisions	45 (20)	I never want to be in a situation where I face disciplinary and/or legal consequences for reporting a patient who is miscarrying (spontaneous or induced), and with the current climate, I genuinely fear that we may be moving toward the criminalization of abortion in many places. That risk is not worth it to me when I could train in so many other places.	78 (25)	A politician is unable to grasp the grey areas of obstetric care and the heartbreaking scenarios we encounter. It is bad enough that hospital administrators police our obstetric practice; we do not need another non-medical person telling us how to practice evidence-based medicine.
Challenges of providing any reproductive care to patients with an abortion ban	41 (19)	I was previously set on Ob-Gyn, but I am now looking seriously at other fields because of the politics surrounding women's health care. I don't want to have to worry about legal repercussions for providing the best care to my patients. This has strongly turned me away from Ob-Gyn as a medical specialty.	65 (21)	As an abortion provider, I know that as much as I care about serving a population with unmet needs, the inevitability of burnout working in a place where abortion is severely limited would be too much.
Challenges of providing patient care that is not reproductive in nature	15 (7)	It will significantly impact the ability of every physician to provide care to their patients, regardless of their specialty, as many medical conditions are exacerbated by pregnancy status	47 (15)	I'm a dermatologist, and this affects our practice, too! We prescribe Accutane every day, and if a patient does become pregnant while on this drug due to contraceptive failure, we recommend termination. We prescribe lots of other teratogenic drugs as well for many different cutaneous diseases, especially methotrexate. I don't know how I can practice in a state where pharmacists might refuse to fill MTX.

Table 3. Practice location decisions that are workforce-related emerging from the inductive analysis of the following: for students applying to residency, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your residency application and ranking decisions below," and for fellows and practicing physicians, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your decision about your job or fellowship programs" among respondents (n=524).

Theme	Students (n=219), n (%)	Example quote	Physicians (n=305), n (%)	Example quote
Not choosing to practice or train in a state with abortion restrictions	77 (35)	This decision has heavily affected my residency application process. Amazing programs that I've highly considered are now at the bottom of my list.	172 (56)	Just finished residency and specifically did not even consider jobs in states that ban the full spectrum of reproductive healthcare or states that looked like they would consider a ban. Overturning of Roe made me basically have to ignore half the country during my search. But given the job market today, finding a position in a state that allows me to actually care for my patients wasn't hard.
Personal belief that an abortion ban is a human rights/body autonomy violation	63 (29)	States that do not respect basic human rights are not places I wish to live or raise a family.	134 (44)	The overturning of Roe is the overturning of basic freedoms, the right to privacy, and bodily autonomy. It is the first step in overturning other rights. It is removing science from medicine. It threatens all doctors whether they provide abortion care or not. I'm likely to leave medicine, then practice in that environment and take those risks.
Access to training and education in abortion	43 (20)	I want to be part of a program where abortion training is easily accessible, and I will not have to go out of state to get this training. I also want to protect these rights for myself and my future patients.	15 (5)	One of my biggest decisions in choosing my state of residency was to allow me every opportunity to learn about women's care at all levels. The overturning will prevent students and residents from reaching their full potential of learning care for women. It is truly unfortunate that men outside of the walls of understanding of medical knowledge think they have the authority to control not only women's bodies but also the education of those to be able to treat women in emergency settings safely and holistically.
Personal or family access to abortion care or family building	36 (16)	I'm a guy, but what about my daughters in the future? What about a pregnancy complication with my wife? What about my patients? This is the problem when people claim moral high ground on the basis of their religion and are placed into positions of power; you end up with a sort of theocracy.	58 (19)	I was planning on looking for underserved community jobs in Idaho, but now that they have an early abortion ban, I will not be. I am actively trying to get pregnant and won't risk my life to pursue a job.
Geographic ties to states with abortion restrictions limiting relocation	18 (8)	I attend medical school in my home state, which hasn't banned abortion as of yet but might do so in the future. If abortion is banned here, I'll likely still rank in-state programs due to the proximity of my family, but I will not rank out-of-state programs where abortion is banned.	38 (12)	Unfortunately, my answers are influenced by the fact that I live in a state with some of the most restrictive policies and have no ability to move. I cannot simply uproot my life to another state due to my feelings on abortion access. I work here, and my husband works here. My family is here. His family is here. The best I can do is to advocate for change, but I must remain in place as the primary breadwinner in my family.
Challenges recruiting to states with abortion restrictions	0 (0)		18 (6)	I'm a program director and am concerned about how this will affect recruiting talented and eager physicians to our state. Our patients already have difficulty accessing the medical system, so if this decision leads to physicians leaving the state, it will only amplify disparities.

Patient Factors Influencing Decisions About Practice Location

Patient Access to Abortion or Full-Spectrum Reproductive Care

Many physicians and medical student respondents want patients to have access to safe and legal abortion. Respondents specifically highlighted concerns that adolescents, underrepresented minority groups, people in rural communities, and lower-income patients would increasingly face challenges in finding abortion providers, exacerbating health disparities (Table 2).

Physicians also noted that restrictions interfere with their ability to provide or refer patients for abortion care. For example, one stated, "I won't practice in a state that limits my ability to provide or refer my patients for care that is safe and necessary for their health and well-being."

Challenges of Providing Reproductive Care to Patients During an Abortion Ban

Reproductive health care providers anticipate moral distress if they are unable to provide abortion care in circumstances like lethal fetal anomalies or pregnancies resulting from rape or incest. An OBGYN physician explained, “Abortion care and prenatal care go hand in hand. This is a field with a lot of gray areas, and elimination of options will harm those who can get pregnant.” Many physicians feared legal repercussions and were disappointed by a perceived lack of institutional support for evidence-based health care.

Do Not Want Politics to Interfere With Medical Care Decisions

Some respondents expressed concern that lawmakers are interfering with medical care. Others emphasized the role physicians play in advocacy and supporting elected officials in favor of essential reproductive health care. A participant stated, “The government should have no standing in a medical decision between physician and patient.”

Challenges of Providing Patient Care That Is Not Reproductive in Nature

Physicians across various fields were concerned that abortion restrictions would adversely impact their clinical practice. For example, a pediatrician noted, “Working with fetal cardiac patients, it is imperative that my patients have access to abortion services if that’s the choice they make that’s best for their families.” In addition, an oncologist worried about restrictions on chemotherapy regimens, a dermatologist had questions about prescribing common medications (like Accutane) that are teratogenic, and a rheumatologist had concerns about prescribing methotrexate.

Workforce-Related Practice Location Decisions

Choosing Not to Practice or Train in a State With Abortion Restrictions

Many respondents living in states with abortion protections stated that they would be unwilling to move to a state with abortion restrictions (Table 3). Others living in restrictive states intend to move or preferentially apply to and rank training programs in states without abortion bans. Trainees described how these decisions compound their stress regarding the highly competitive match process. Some still felt pressured to apply everywhere, regardless of their personal preferences, stating, “Residency is already so competitive, so unfortunately, I feel like I have to apply everywhere, but I would definitely preferentially rank somewhere that I would have access to abortion care and that my patients would as well.”

Challenges Recruiting to States With Abortion Restrictions

Some residency and fellowship program directors and administrative leadership in states with restrictive abortion laws are concerned about recruiting and retaining residents, fellows, and faculty. Many foresee the reluctance of trainees and faculty to work in restrictive states. A program leader said, “I am an APD at an academic medical center in the Midwest. I have already been told by two residents that they had planned to stay in the state to practice but are now leaving solely because of the lack of reproductive rights in our state. I fear we will rapidly lose amazing physicians.”

Personal Belief That Abortion Restrictions Violate Human Rights and/or Bodily Autonomy

A substantial portion of respondents described the overturning of *Roe v Wade* as a human rights violation and criticized its negative impact on patients’ bodily autonomy. Others discussed the potential moral injury from practicing in a state whose laws and policies prevent clinicians from providing evidence-based medical care.

Respondents connected states’ abortion policies to their overarching sociopolitical climates, noting that bans and restrictions may portend other harmful (eg, racist, homophobic, transphobic) policies. A medical student said, “Extremely cautious about applying to these states who have denied abortion care. Not only because of abortion care but also because these states are notoriously anti-LGBTQ+ and hold racist values. I do not want to live and work and raise a family in that environment, where I am not respected and have less human rights than others.”

Access to Abortion Training and Education

Students applying to OBGYN and family medicine expressed that their application decisions would be shaped by access to proper training in abortion care. Applicants to residency and fellowship recognize that selecting programs in abortion-restricted states may limit access to adequate training. Multiple students noted that they intend to inquire about abortion training during the residency application process.

Some recognized that trainees in abortion-restricted states could seek abortion training out-of-state. For example, a respondent said, “I plan to first rank programs in states with full spectrum reproductive health access, followed by programs that are intentional about providing training for their residents with full support (financial, housing, etc) to leave the state for abortion training.” However, current trainees also discussed challenges in obtaining abortion training, including professional, administrative, and financial barriers.

Geographic Ties to States With Abortion Restrictions Limiting Relocation

Some noted that geographic relocation is a privilege not afforded to everyone equally. The decision to move is often influenced by distance to a support network, job benefits for the respondent or their spouse, housing, and childcare. Such geographic ties discourage or prevent many medical students and physicians from leaving their state of residence despite their personal or professional opposition to abortion restrictions.

Some said they understand the risks of staying in a state with abortion restrictions. If necessary, they would travel out of state to receive an abortion, again recognizing their mobility is a privilege. A respondent said, “I definitely would prefer to be in a state that maintains access to abortion. Unfortunately, those are not states where my family lives, and I am grateful that I

have enough privilege if I needed an abortion, I could leave the state."

Personal or Family Access to Abortion Care or Family Building

Respondents were concerned about practicing in a location that limits their options for receiving comprehensive reproductive health care. Multiple respondents highlighted that they did not want to be forced to carry a pregnancy if they could not get an abortion, especially during training. Others specifically cited medical conditions that would make pregnancy physically challenging and even contraindicated as a reason to ensure they had access to abortion care. A respondent said, "I am a medical student with chronic conditions that make pregnancy life-threatening for me. Although I am on contraceptives, nothing is 100%, and I want to be able to protect my life and well-being in case I do accidentally get pregnant."

In addition, physicians with infertility undergoing in vitro fertilization noted that practicing in a state where life is defined as beginning at fertilization would make family building significantly more challenging. Commonly, respondents stated they were concerned about care for themselves, their children, or their partners, underscoring the importance of recognizing that physicians, too, need access to care.

Additional Themes

No Impact

Few medical students and physicians stated that the *Dobbs* decision would not impact their choice of practice location (Table 4). Some indicated that the residency and fellowship match were too competitive to make decisions based on abortion legislation. For example, those who apply to every program in their field may end up applying to programs in states with abortion restrictions to increase their likelihood of matching.

Table 4. Practice location decisions that are workforce-related emerging from the inductive analysis of the following: for students applying to residency, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your residency application and ranking decisions below," and for fellows and practicing physicians, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your decision about your job or fellowship programs" among respondents (n=524).

Theme	Students (n=219), n (%)	Example quote	Physicians (n=305), n (%)	Example quote
No impact	30 (14)	Matching and getting into a program is challenging enough considering the various factors at play; this decision will not be part of deciding which states or programs I end up applying to.	11 (4)	It will have zero impact on my decisions regarding jobs/fellowships.
Expressed support for overturning *Roe v Wade*	17 (8)	The overturning of Roe v. Wade is long overdue. It was not right in the first place, as the Supreme Court made clear in its ruling. Babies deserve to live inside and outside the womb.	18 (6)	I am supportive of the overturn and believe it will be better for our patients and medical care to ban an inhumane practice like abortion. Human lives in the womb deserve protection just like all of our other patients at any age and ability.

Expressed Antiabortion Sentiment and/or Support for Overturning Roe v Wade

Physicians and medical students who expressed antiabortion ("pro-life") views supported the Supreme Court decision (Table 4). Multiple respondents noted that they would purposefully seek out practice environments where abortion restrictions existed. Reasons for this include not supporting abortion care for any indication, stating that they do not view abortion as health care, a desire to "preserve life," and a desire to "protect the unborn." Multiple respondents discussed that abortion is an issue that should be legislated at the state level.

Discussion

Principal Findings

Our study shows that abortion restrictions will have a substantial impact on the physician workforce in patient care and practice location decisions. The 3 most common themes were patient access to care, not choosing to practice or train in a state with abortion restrictions, and personal belief that an abortion ban is a human rights/body autonomy violation. This study enhances emerging literature about the impacts of abortion restrictions on the physician workforce, including physicians and medical students at all levels of training across all 50 states within both reproductive and nonreproductive health fields.

Respondents shared concerns that abortion restrictions will negatively impact their ability to provide high-quality, comprehensive reproductive care. This was evident among trainees who provide abortion care, like OBGYN residents, who expressed concerns about new or worsening barriers to obtaining foundational skills like first-trimester uterine aspiration at their primary institution [23,24]. OBGYN trainees also cited multiple barriers to obtaining foundational abortion care skills at their primary institutions. Some programs have created away rotation opportunities for residents unable to obtain comprehensive abortion training at their own institutions [25]. However, there are many barriers to these programs, including obtaining state-based medical licenses, getting funding and organizational affiliations in place, and disruptions to families when living in another state [25].

Even within nonreproductive health care fields, respondents shared concerns about the downstream effects of abortion restrictions on clinical training and practice. In the 2 weeks following the *Dobbs* decision, only 38.5% of a list of 187 societies across a wide variety of specialties had made a statement about the decision [26]. Respondents from specialties

that do not provide abortion care noted concern for restricted use of potentially abortifacient or teratogenic medications and worsening health among patients whose physical or mental health will be adversely impacted by restrictions.

Physicians and medical students also worried that abortion restrictions would deleteriously affect their personal health and well-being. Restrictions hold significant health implications for reproductive-age women, a large and growing demographic of the physician workforce [27]. Recent studies have reported that abortion is common among physicians, affirming that physicians, too, need safe, legal access to abortion [28]. This study informs future medical education and occupational health research by elevating trainees' and employees' concerns. As highlighted by the medical student responses on geographic ties and competitiveness of the match process, it is critical to recognize the multifactorial decision-making involved in where to complete residency training. While before the *Dobbs* decision, telehealth may have been able to bridge the gaps in access to abortion care, this is less likely to be possible in the current landscape [29]. Medical schools and hospitals, especially those in restrictive states, must recognize this and prepare to navigate the adverse health, financial, and legal repercussions their employees may face. Otherwise, disparate abortion access may increase health disparities within the physician workforce and threaten its diversity and resiliency [30].

If medical students do not want to practice in states with abortion restrictions in place, it is less likely that they will establish their practice in those locations. In 2022, 55.2% of those completing training established their practice in the same state where they completed residency [31]. The lack of physicians who are willing to practice in states with abortion restrictions can further poor health outcomes in maternity care deserts [32,33]. Idaho is a notable example, where 41% of OBGYN physicians consider leaving and cite restrictive abortion laws as a motivation [34,35]. Idaho has the lowest rate of physicians per 100,000 people in the entire country [34,35].

Some physicians stated that abortion restrictions would not impact them or that they support them. Notably, a subset of "no impact" responders shared that the scarcity of available positions, particularly within highly competitive specialties and for historically marginalized applicants, outweighs their personal opposition to abortion restrictions. Others acknowledged the futility of setting preferences since the match is ultimately complex and multifactorial.

Limitations

This study may be limited by self-selection bias, given its recruitment of medical students and practicing physicians on social media. Of the respondents who did not prefer to apply where abortion access was preserved, a substantial number (41.5%) provided a free response, indicating that we had a spectrum of views on abortion rights. Furthermore, this sample is focused on physicians and does not represent other health care workforce members who are likely also impacted by abortion restrictions.

Conclusion

The findings of this study captured responses to abortion restrictions before the 2023 Match cycle and provided context to the recent AAMC data showing that residency applications disproportionately decreased in restrictive states [13]. Narrative responses bolster our original quantitative data, affirming that access to full-spectrum reproductive health care was highly valued personally and professionally by most physicians [15].

This study shows that abortion restrictions are having an impact on the practice location preferences of the physician workforce due to both patient care and personal factors. It is important that state policy makers and others who are considering abortion restrictions also consider how to address these concerns of physicians and medical students, to avoid worsening geographic maldistribution of physicians and worsening access to care from physicians for their citizens.

Acknowledgments

Generative artificial intelligence was not used to aid in writing this manuscript.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly to protect the anonymity of participants due to the sensitive subject matter of this manuscript but are available from the corresponding author on reasonable request.

Authors' Contributions

MSL and SAB contributed equally as cofirst authors. SAB and MSL had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. SAB, MSL, JAG, and VMA contributed to the concept and design. All authors contributed to the acquisition, analysis, or interpretation of data. SAB, MSL, SMM, JAG, and VMA contributed to the drafting of the manuscript. All authors contributed to the critical revision of the manuscript for important intellectual content. MSL performed statistical analysis. JAG and VMA contributed to administrative, technical, or material support. JAG and VMA performed supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[PDF File (Adobe PDF File), 316 KB - ijmr_v14i1e55035_app1.pdf]

References

1. Nash E, Guarnieri I. Six months post-Roe, 24 US States have banned abortion or are likely to do so: a roundup. Guttmacher Institute. 2023. URL: https://www.guttmacher.org/node/300254/printable/print [accessed 2023-03-13]
2. Abortion laws by state. Center for Reproductive Rights. 2022. URL: https://reproductiverights.org/maps/abortion-laws-by-state/ [accessed 2023-03-13]
3. Sabbath EL, McKetchnie SM, Arora KS, Buchbinder M. US obstetrician-gynecologists' perceived impacts of post-Dobbs v Jackson state abortion bans. JAMA Netw Open 2024;7(1):e2352109 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.52109] [Medline: 38231510]
4. Keegan G, Francis M, Chalmers K, Hoofnagle M, Noory M, Essig R, et al. Trauma of abortion restrictions and forced pregnancy: urgent implications for acute care surgeons. Trauma Surg Acute Care Open 2023;8(1):e001067 [FREE Full text] [doi: 10.1136/tsaco-2022-001067] [Medline: 36744294]
5. Stevenson AJ, Root L. Trends in maternal death post-Dobbs v Jackson Women's Health. JAMA Netw Open 2024;7(8):e2430035 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.30035] [Medline: 39190311]
6. Singh P, Gallo MF. National trends in infant mortality in the US after Dobbs. JAMA Pediatr 2024:e244276. [doi: 10.1001/jamapediatrics.2024.4276] [Medline: 39432283]
7. Blair JE, Mayer AP, Caubet SL, Norby SM, O'Connor MI, Hayes SN. Pregnancy and parental leave during graduate medical education. Acad Med 2016;91(7):972-978. [doi: 10.1097/ACM.0000000000001006] [Medline: 26606722]
8. Stentz NC, Griffith KA, Perkins E, Jones RD, Jagsi R. Fertility and childbearing among American female physicians. J Womens Health (Larchmt) 2016;25(10):1059-1065 [FREE Full text] [doi: 10.1089/jwh.2015.5638] [Medline: 27347614]
9. Feinberg EC, Kawwass JF, Cedars MI. Roe v Wade and the threat to fertility care. Obstet Gynecol 2022;140(4):557-559. [doi: 10.1097/AOG.0000000000004928] [Medline: 35861359]
10. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. International Journal of Qualitative Methods 2017;16(1):160940691773384. [doi: 10.1177/1609406917733847]
11. Desai A, Holliday R, Wallis M, Thornewill A, Bahraini NH, Borges LM. Policy changes as a context for moral injury development in the wake of Dobbs v Jackson Women's Health Organization. Obstet Gynecol 2023;141(1):15-21. [doi: 10.1097/AOG.0000000000005009] [Medline: 36701606]
12. Martin LA, Debbink M, Hassinger J, Youatt E, Harris LH. Abortion providers, stigma and professional quality of life. Contraception 2014;90(6):581-587. [doi: 10.1016/j.contraception.2014.07.011] [Medline: 25131444]
13. Training location preferences of U.S. medical school graduates post Dobbs v Jackson Women's Health Organization decision. American Association of Medical Colleges. URL: https://www.aamc.org/advocacy-policy/aamc-research-and-action-institute/training-location-preferences [accessed 2023-04-16]
14. Hammoud MM, Morgan HK, George K, Ollendorff AT, Dalrymple JL, Dunleavy D, et al. Trends in obstetrics and gynecology residency applications in the year after abortion access changes. JAMA Netw Open 2024;7(2):e2355017 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.55017] [Medline: 38324311]
15. Bernstein SA, Levy MS, McNeilly S, Fishbach S, Jain S, Gold JA, et al. Practice location preferences in response to state abortion restrictions among physicians and trainees on social media. J Gen Intern Med 2023;38(10):2419-2423. [doi: 10.1007/s11606-023-08096-5] [Medline: 36823418]
16. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
17. O'Cathain A, Thomas KJ. "Any other comments?" open questions on questionnaires - a bane or a bonus to research? BMC Med Res Methodol 2004;4:25 [FREE Full text] [doi: 10.1186/1471-2288-4-25] [Medline: 15533249]
18. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
19. McCann A, Walker AS, Sasani A, Johnston T. Tracking the states where abortion is now banned. New York Times. URL: https://www.nytimes.com/interactive/2024/us/abortion-laws-roe-v-wade.html [accessed 2024-11-28]
20. Diversity in medicine: facts and figures 2019. American Association of Medical Colleges. URL: https://www.aamc.org/data-reports/workforce/report/diversity-medicine-facts-and-figures-2019 [accessed 2023-04-09]
21. ACGME Data Resource Book. Accreditation Council for Graduate Medical Education. URL: https://www.acgme.org/about-us/publications-and-resources/graduate-medical-education-data-resource-book/ [accessed 2023-04-09]
22. Women's reproductive health. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/reproductive-health/women-health/?CDC_AAref_Val=https://www.cdc.gov/reproductivehealth/womensrh/index.htm [accessed 2023-04-09]
23. Steinauer JE, Turk JK, Pomerantz T, Simonson K, Learman LA, Landy U. Abortion training in US obstetrics and gynecology residency programs. Am J Obstet Gynecol 2018;219(1):86.e1-86.e6. [doi: 10.1016/j.ajog.2018.04.011] [Medline: 29655963]

24. Vinekar K, Karlapudi A, Nathan L, Turk JK, Rible R, Steinauer J. Projected implications of overturning Roe v Wade on abortion training in U.S. obstetrics and gynecology residency programs. Obstet Gynecol 2022;140(2):146-149. [doi: 10.1097/AOG.0000000000004832] [Medline: 35852261]
25. !!ATTN #OBGYN Residents:!! OHSU is committed to educating the next generation of #abortion providers. Residents with limited or no access to abortion training are invited to apply a Complex Family Planning elective at OHSU for the 2023-2024 academic year. @eeyanmiller. 2023 Mar 15. URL: https://twitter.com/eeyanmiller/status/1636103934434758658 [accessed 2023-04-09]
26. Levy MS, Fishbach S, Salles A, Arora VM. When do medical societies take a stand on abortion? J Gen Intern Med 2023;38(6):1559-1562 [FREE Full text] [doi: 10.1007/s11606-022-07867-w] [Medline: 36323828]
27. Physician specialty data report. American Association of Medical Colleges. URL: https://www.aamc.org/data-reports/workforce/report/physician-specialty-data-report [accessed 2023-04-09]
28. Levy MS, Arora VM, Talib H, Jeelani R, Duke CMP, Salles A. Abortion among physicians. Obstet Gynecol 2022;139(5):910-912. [doi: 10.1097/AOG.0000000000004724] [Medline: 35576350]
29. Koenig LR, Becker A, Ko J, Upadhyay UD. The role of telehealth in promoting equitable abortion access in the United States: spatial analysis. JMIR Public Health Surveill 2023;9:e45671. [doi: 10.2196/45671] [Medline: 37934583]
30. Nobel K, Luke AA, Rice WS. Racial disparities in pregnancy options counseling and referral in the US South. Health Serv Res 2023;58(1):9-18 [FREE Full text] [doi: 10.1111/1475-6773.14049] [Medline: 36068681]
31. Table C6. physician retention in state of residency training, by state. American Association of Medical Colleges. URL: https://www.aamc.org/data-reports/students-residents/data/report-residents/2020/table-c6-physician-retention-state-residency-training-state [accessed 2023-04-09]
32. DiPietro Mager NA, Zollinger TW, Turman JE, Zhang J, Dixon BE. Routine healthcare utilization among reproductive-age women residing in a rural maternity care desert. J Community Health 2021;46(1):108-116. [doi: 10.1007/s10900-020-00852-6] [Medline: 32488525]
33. Wallace M, Dyer L, Felker-Kantor E, Benno J, Vilda D, Harville E, et al. Maternity care deserts and pregnancy-associated mortality in Louisiana. Womens Health Issues 2021;31(2):122-129 [FREE Full text] [doi: 10.1016/j.whi.2020.09.004] [Medline: 33069560]
34. Cunningham M. Survey shows Idaho's maternal health doctors are leaving the state, or soon will. Idaho Capital Sun. 2023 Apr 7. URL: https://idahocapitalsun.com/2023/04/07/survey-shows-idahos-maternal-health-doctors-are-leaving-the-state-or-soon-will/ [accessed 2023-04-09]
35. State physician workforce data report. American Association of Medical Colleges. URL: https://store.aamc.org/downloadable/download/sample/sample_id/506/ [accessed 2024-11-28]

Abbreviations

AAMC: American Association of Medical Colleges
OBGYN: obstetrics and gynecology
TRAP: Targeted Regulation of Abortion Providers

Edited by T de Azevedo Cardoso; submitted 05.12.23; peer-reviewed by A Arbabisarjou, A Stephenson-Famy; comments to author 03.07.24; revised version received 25.08.24; accepted 22.10.24; published 07.01.25.

Please cite as:
Levy MS, Bernstein SA, McNeilly SM, Liberty A, Fishbach S, Jain S, Gold JA, Arora VM
Student and Physician Views of How the Dobbs Decision Affects Training and Practice Location Preferences: Cross-Sectional Questionnaire Study
Interact J Med Res 2025;14:e55035
URL: https://www.i-jmr.org/2025/1/e55035
doi:10.2196/55035
PMID:

Original Paper

Exploring Barriers to Patients' Progression in the Cardiac Rehabilitation Journey From Health Care Providers' Perspectives: Qualitative Study

Shri Harini Ramesh[1], MAS; Darwin Jull[1], BEng; Hélène Fournier[2], PhD; Fateme Rajabiyazdi[1,3], PEng, MSc, PhD

[1]Department of Systems and Computer Engineering, Faculty of Engineering and Design, Carleton University, Ottawa, ON, Canada

[2]National Research Council Canada, Moncton, NB, Canada

[3]Bruyère Research Institute, Bruyère, Ottawa, ON, Canada

Corresponding Author:
Fateme Rajabiyazdi, PEng, MSc, PhD
Department of Systems and Computer Engineering
Faculty of Engineering and Design
Carleton University
1125 Colonel By Dr
Ottawa, ON, K1S 5B6
Canada
Phone: 1 613 520 2600 ext 8229
Email: fateme.rajabiyazdi@carleton.ca

Abstract

Background: Cardiovascular diseases are one of the leading causes of mortality globally. Cardiac rehabilitation (CR) programs are crucial for patients recovering from cardiac events, as they help reduce the risk of recurrent events and support patient recovery. The patient's journey in CR spans the stages before, during, and after the program. Patients have to progress through each stage of CR programs successfully to complete the entire CR journey and get the full benefits of CR programs, but numerous barriers within this journey can hinder patient progression.

Objective: This study aims to explore the barriers to progression at all stages of the CR patient journey from the perspectives of health care providers involved in CR care.

Methods: This qualitative study involved semistructured interviews with health care providers involved in CR care from July 2023 to January 2024. A purposive maximal variation sampling method was used to target providers with diverse demographics and specialties. Snowball sampling was used to recruit participants, leveraging the existing networks of participants. Each interview lasted between 30 and 45 minutes. Interviews were recorded, transcribed verbatim, and analyzed using an inductive thematic analysis approach. Data analysis was conducted from August 2023 to February 2024.

Results: Ten health care providers, comprising 7 females and 3 males, were interviewed. Their roles included physician, program director, nurse manager, clinical manager, nurse coordinator, nurse, physiotherapist, and kinesiologist. The analysis identified four overarching themes related to barriers to progression in the CR journey: (1) patients not being referred to CR programs, (2) patients not enrolling in CR programs, (3) patients dropping out of CR programs, and (4) patients' lack of adherence to lifestyle changes post-CR programs.

Conclusions: In light of the growing interest in technological interventions in CR programs, we proposed 4 potential technological solutions to address the barriers to progression identified in our analysis. These solutions aim to provide a foundation for future research to guide the development of effective technologies and enhance patient progression within the CR journey.

(Interact J Med Res 2025;14:e66164) doi:10.2196/66164

KEYWORDS

cardiac rehabilitation; health care providers; CR patient journey; qualitative study; barriers; technology

Introduction

Cardiovascular diseases are one of the leading causes of mortality globally [1]. Cardiac rehabilitation (CR) programs are often recommended for patients who have experienced a cardiac event to reduce the risk of recurrent cardiovascular events and support patient recovery [2]. CR programs are comprehensive and multidisciplinary, encompassing exercise training, risk factor management, health education, and psychosocial support [3,4]. Research has shown that participation in CR programs significantly reduces morbidity and mortality associated with cardiac events [5,6] while also enhancing patients' health-related quality of life and psychological well-being [7,8].

The patient's CR journey begins even before enrollment and continues through maintaining lifestyle changes after completing the program [9,10]. It begins with a provider's referral to the CR program, made before discharge from the hospital following a cardiac event or surgery [11]. The journey then involves enrolling in, participating in, and completing the CR program [12]. After completing the program, the journey continues with integrating and maintaining the lifestyle changes and health habits learned throughout the program [13]. Patients must successfully progress through each stage to complete the entire CR journey in order to get the full benefits of CR programs. However, numerous barriers exist within the CR patient journey that could affect progression.

Previous research has identified various barriers in CR, such as low patient interest, comorbidities, travel complications, limited program availability, scheduling conflicts, and financial constraints [14-18]. However, these studies focused only on specific stages of the CR patient journey, such as referral, enrollment, completion, or postcompletion, but not throughout the entire CR patient journey. Understanding the barriers to progression throughout the entire CR patient journey is essential because each stage of the journey is interconnected, and barriers at one stage can influence subsequent stages [19,20]. Moreover, existing studies predominantly focus on the patient's perspective. It is equally important to consider the perspectives of the providers who deliver CR care, as they encounter a diverse range of patients with varying needs. Insights from those who deliver care can provide valuable information on improving CR delivery and addressing barriers effectively [21].

The aim of this study was to explore the barriers to progression throughout the CR patient journey from the perspectives of health care providers involved in CR care. We conducted semistructured interviews with 10 providers and identified 4 barriers to progression within the CR patient journey. In light of the growing interest in technological interventions in CR programs [20,22], we proposed 4 potential technological solutions to overcome the barriers to progression identified in our study. The novelty of our work lies in the identification of barriers to progression from the perspectives of health care providers and the proposal of potential solutions to address these challenges. We hope these potential solutions provide the foundation for future research and guide the development of effective technologies to overcome barriers to progression within the CR patient journey.

Methods

Ethical Considerations

This study protocol was approved by the ethics committees of Carleton University (clearance number 119779) and the National Research Council Canada (clearance number 2023-108). Ethical review was conducted per institutional guidelines, with no exemptions or waivers applied. Written informed consent was obtained before each interview. Participants were informed of their right to withdraw at any time without consequence. No secondary data were used. All data were anonymized to ensure privacy and confidentiality. No compensation was provided to the participants.

Overview

This qualitative study is reported in accordance with the COREQ (COnsolidated criteria for REporting Qualitative research) guidelines [23] (Multimedia Appendix 1). We conducted semistructured interviews to explore the barriers to progression in the CR patient journey from the perspective of the health care providers involved in CR care. We chose interviews to ensure that participants could freely share their thoughts and experiences. The semistructured format allowed us to ask follow-up questions in order to gain an in-depth understanding of their perspectives. In addition, the confidential nature of the interviews encouraged open and honest sharing [24]. Eligible participants included health care providers who refer patients to CR, those who provide CR services, and those who manage CR programs. Eligible participants were also required to be proficient in English. Exclusion criteria included individuals not involved in direct patient care or CR, such as administrative staff. To ensure that the perspectives of a diverse group of participants from various specialties in CR care were included, we used purposive sampling [25]. To further expand the participant pool, we used snowball sampling methods [26]. HF, who has extensive connections with health care providers involved in CR care and CR centers, shared the study objectives and recruitment details within her network. Through this process, she received the email addresses of potential participants. Subsequently, SHR contacted these potential participants via email, providing detailed information about the study and inviting them to participate. We contacted 11 potential participants, of whom 10 agreed to participate, while 1 did not respond. No participants dropped out after providing informed consent. No prior relationship was established between the researcher and the participants before the study commenced, and participants knew only that the researcher was conducting the study as part of an academic project focused on CR.

Data Collection

Data collection took place from July 2023 to January 2024. The interviews were conducted virtually via Microsoft Teams, with each session lasting between 30 and 45 minutes. No repeat interviews were carried out. Each participant was interviewed once during the study. The researcher (SHR), a female PhD student familiar with the subject matter and trained in qualitative

research, conducted the interviews, asked questions and follow-up questions as needed, and took notes during the sessions. While conducting the interviews, the researcher maintained a neutral stance, ensuring that her personal views or assumptions did not influence the questions asked or the participants' responses. The research team collaboratively and iteratively developed an interview script and questions (Multimedia Appendix 2). The questions were open-ended to allow for an in-depth exploration of the providers' perspectives on the barriers to progression in the CR patient journey, with follow-up questions asked when appropriate. We conducted a pilot test of the interview guide with a health researcher to ensure clarity, relevance, and logical flow of the questions. Feedback from this pilot test was used to refine the questions and the interview guide before data collection commenced.

Data Analysis

Our analysis was conducted from August 2023 to February 2024 inductively and did not aim to validate any preexisting theories or hypotheses. We followed a thematic analysis approach following the 6 steps outlined by Braun and Clarke [27]. Two analysts, SHR and a male undergraduate student (DJ) with subject knowledge and qualitative research training, were responsible for the analysis. Disagreements in coding and theme assignment were resolved through consensus facilitated by senior female researchers FR and HF, both of whom are familiar with the subject matter and hold PhDs with extensive experience in qualitative research. We adopted a reflexive approach, acknowledging the influence of researchers' views and experiences [28]. In step 1, both analysts read interview transcripts and took notes. In step 2, they reviewed the first 2 transcripts line by line and identified prominent codes to generate a codebook using NVivo software (version 14.23; Lumivero), which was then applied to subsequent interviews. Codes were iteratively refined every 2 interviews. Step 3 involved systematically grouping codes into candidate themes. In step 4, these themes were compared and refined. In step 5, other team members contributed to further theme refinement. Thematic saturation was assessed using a saturation table that documented the themes found in each interview. Finally, in step 6, we produced a report summarizing our findings. Transcripts were not returned to participants for comment or correction. The analysis was conducted based solely on the interview data collected during the sessions.

Rigor and Trustworthiness

Rigor and trustworthiness were achieved by ensuring credibility, transferability, dependability, and confirmability throughout the research process [29]. To address credibility, we conducted peer debriefing sessions within our research team, which provided an external check on our interpretation of the data, analysis, and coding decisions. Dependability was achieved through detailed documentation of themes and subthemes, providing transparency, and enabling replicability of the coding process. To support transferability, we provided comprehensive descriptions of our study design, participant selection criteria, and data collection methods, allowing others to evaluate the applicability of our findings to various CR settings. As Lincoln and Guba [30] highlighted, confirmability is achieved when credibility, transferability, and dependability are established. We ensured confirmability by documenting the rationale for theoretical, methodological, and analytical decisions throughout the study.

Results

Overview

Ten health care providers, comprising 7 females and 3 males involved in CR care, including a physician, a program director, a clinical manager, a nurse manager, a nurse coordinator, a nurse, 3 physiotherapists, and a kinesiologist, were interviewed (Table 1). The thematic analysis revealed four overarching themes concerning barriers to progression within the CR patient journey, as perceived by these providers: (1) patients not being referred to CR programs, (2) patients not enrolling in CR programs, (3) patients dropping out of CR programs, and (4) patients' lack of adherence to lifestyle changes post-CR programs. Multimedia Appendix 3 provides illustrative quotes within each of the themes identified. Thematic saturation was reached after analyzing 7 interviews [31] (Multimedia Appendix 4). This was determined using a saturation table, where no new themes emerged after the seventh interview. To ensure robustness, 3 additional interviews were conducted, which confirmed the saturation of themes. The early achievement of saturation reflects the shared perspectives of participants, despite their varied roles within CR care.

Table 1. Characteristics of interviewed participants.

Characteristics	Values
Total, n	10
Sex, n (%)	
Female	7 (70)
Male	3 (30)
Specialty, n (%)	
Clinical manager	1 (10)
Nurse manager	1 (10)
Nurse	1 (10)
Physician	1 (10)
CR[a] program director	1 (10)
Kinesiologist	1 (10)
Physiotherapist	3 (10)
Nurse coordinator	1 (10)
Practice setting, n (%)	
Urban	7 (70)
Urban and rural	2 (20)
Rural	1 (10)
Age (year), mean (SD)	47.9 (10.2)
Years of experience, mean (SD)	18.7 (9.5)

[a]CR: cardiac rehabilitation.

Patients Not Being Referred to CR Programs

One of the barriers to progression within the CR patient journey is patients not being referred to CR programs, as without referrals, patients may not have access to CR programs (Table S1 in Multimedia Appendix 5). Unintentional biases in referrals, such as those related to health conditions, geography, age, and gender, can contribute to patients not being referred to CR programs. As one provider noted, “There is a lot of inherent bias in terms of who is being referred to rehab” [P3]. Providers may hesitate to refer patients with comorbidities or complex health conditions to CR programs due to concerns about their ability to participate. In addition, if the patient is not residing near a CR center, the provider may assume that it is difficult for the patient to travel and, as a result, be hesitant to refer the patient to a CR program. Providers may perceive that older adults might be frail or disconcerted after a cardiac event or surgery, which may result in providers not referring these patients to a CR program. Some providers may be hesitant to refer certain patients to CR programs, considering their gender-specific responsibilities, such as work, family, and parental roles, which could contribute to women being overlooked for referrals.

Furthermore, another reason for not referring is the providers’ heavy workloads. The paperwork involved in referring patients to CR programs can be time-consuming, which may discourage some providers from making these referrals. Although automatic referral systems are available in some CR centers, providers’ endorsements are crucial for motivating patients; as one provider emphasized, “Endorsements from [providers] are key for [motivating patients]” [P1]. However, individually discussing the benefits of CR with every patient is time-consuming and requires much effort. Referring providers may also lack awareness of the various programs available at the CR center. As noted, “[Referring providers] just don’t know all the services that are available” [P8]. Consequently, providers may hesitate to refer patients due to uncertainty about the programs available. In addition, providers mentioned that it is important to carefully select a CR program that aligns with the patient’s needs, as some programs with high-impact exercises might be appropriate for immediate needs, whereas low-impact exercise-based programs could offer better long-term benefits. However, identifying the best CR program for each patient during referral can be challenging for providers without access to a wholesome list of CR programs.

Patients Not Enrolling in CR Programs

Another barrier to progression in the CR patient journey is patients not enrolling in CR programs after referral. After patients are referred, they must enroll and start the program to benefit from it. Providers mentioned various factors contributing to this nonenrollment (Table S2 in Multimedia Appendix 5). Providers often observed that a substantial number of patients lack awareness and knowledge about CR, with many patients being unaware of its importance and benefits. Providers also noted that patients sometimes hold misconceptions about CR, believing that they can recover by themselves at home or that

surgical treatment alone is sufficient without rehabilitation. Providers think that inconvenient waiting periods also contribute to patients not enrolling in CR programs. For some patients, the waiting period may be perceived as too short after surgery, leading them to feel unready to start the program. Conversely, for other patients, the waiting period between referral and the program's start may seem excessively long. In addition, CR programs are sometimes offered on fixed schedules, where groups of patients begin the program simultaneously. If patients miss the upcoming enrollment window, they will need to wait for the next scheduled start date, which can lead to delays in enrolling in CR programs.

Providers also think that cultural restrictions can prevent some patients from enrolling in CR programs, particularly when there is discomfort or disallowance in mixed-gender sessions due to cultural or religious beliefs. Providers have mentioned that financial barriers are a concern for some patients and can affect their enrollment. Some patients may not have insurance coverage, or they may be unaware of what services their insurance will cover. Financial barriers can also include the costs associated with transportation and parking. In addition, in some CR centers, patients have the option to participate in CR programs virtually through videoconferencing and web-based meeting platforms such as Zoom (Zoom Video Communications). Providers mentioned a lack of technical knowledge and equipment requirements as factors contributing to nonenrollment in these virtual CR (VCR) programs. They noted that patients may encounter obstacles due to the need for reliable Wi-Fi, laptops, tablets, email addresses, and smartwatches to monitor vitals such as heart rate for VCR, which may not be accessible to all patients.

While providers have mentioned several factors for patients not enrolling in CR programs, they believe that some factors remain hidden and express uncertainty about fully understanding why some patients do not enroll in these programs. Providers have also highlighted that while they try to find these factors through patient data recorded in electronic medical records (EMRs), they are unable to fully explore and uncover the underlying reasons solely from EMRs due to missing or insufficient data. As one provider noted, "We try and capture what we can from [the EMR], so we can count visits and things like that within the EMR. It is definitely not perfect" [P7]. Providers expressed the need for a tracking system to understand potential hidden factors that contribute to patients not enrolling. As one provider mentioned, "[We] do need a little bit of like...something to keep track of, because there's quite a few patients that come through. So just to have...what date they start...it is nice to have an overview of the patient's journey" [P8].

Patients Dropping Out of CR Programs

Dropping out of a CR program before completing it is also a barrier to progression in the CR patient journey, as it prevents patients from fully benefiting from the CR program (Table S3 in Multimedia Appendix 5). Providers observed that reproductive and hormonal conditions can deter female patients from completing or fully participating in CR programs. These conditions include concerns about the safety of exercise during pregnancy, managing pregnancy-related fatigue, menopause issues, and menstrual cycle discomfort, such as severe menstrual cramps. In addition, providers recognized that a lack of support is a factor for dropout, including insufficient family support to manage home obligations such as caregiving for children or older adult family members or a lack of peer support for mental and emotional well-being during the rehabilitation process.

Providers noted that low motivation, low self-efficacy, and a lack of belief in one's ability to perform exercises or improve one's lifestyle contribute to a lack of participation and eventual dropout from the program. As one provider mentioned, "They just don't [participate] because they have no motivation to" [P8]. In addition, providers have observed that low accountability in some patients, requiring an external push and motivation from family or providers, can also lead to dropout. Although providers make efforts to encourage patients by emphasizing the benefits of continuing rehabilitation, some patients may still find it challenging to stay engaged with their rehabilitation, which can contribute to dropouts. As one provider mentioned, "We try to help them understand that if they [continue rehabilitation], their health can get better" [P9].

Patients may face several challenges in reaching CR centers, particularly when they lack access to a car or public transportation or when adverse weather conditions, such as extreme heat or cold, are present. Providers mentioned that these challenges in reaching CR centers can contribute to patient dropouts. In addition, the lack of multilingual support and translators in some CR programs can be challenging for patients who need services in other languages. Providers mentioned that these language barriers can also contribute to dropouts. Providers highlight that some patients may be frail due to medical conditions such as shortness of breath, arthritis, or chronic pain, which can prevent them from performing the required exercises, leading to frustration and dropout. Furthermore, in VCR, patients often need to interact with digital interventions for extended periods. Providers mentioned that this prolonged interaction can be one of the factors contributing to dropouts, particularly in VCR.

Patients' Lack of Adherence to Lifestyle Changes Post-CR Programs

Providers recommend that patients continue following the lifestyle changes they learned during the CR program even after finishing the CR program cycle to achieve better health-related quality of life. However, patients sometimes struggle to adhere to these recommendations, which becomes a barrier to progression in their journey (Table S4 in Multimedia Appendix 5). Providers noted that patients could experience persistent feelings of sadness, hopelessness, stress, and anxiety about daily life and responsibilities, which could lower their motivation and hinder their adherence to lifestyle changes. A lack of personal drive could also contribute to not adhering, as patients may struggle to set and achieve personal health goals independently. Providers noted that financial constraints may impact adherence to post-CR programs, with high costs for healthy food, gym memberships, medications, and inadequate insurance coverage preventing patients from fully following recommended lifestyle changes. Providers try to emphasize the importance of physical exercise in maintaining a healthy

lifestyle, but only a small portion of patients follow this advice. As one provider noted, "The reality is that a small percentage of patients, once they graduate rehab, remain as physically active as they were during rehab" [P3]. In addition, providers mentioned that some patients face challenges in performing physical exercises due to factors such as lack of space at home, exercise-induced pain or discomfort, and a lack of resources to fully understand and perform exercises without pain or discomfort. These challenges lead to frustration and doubts about their ability to perform exercises, as mentioned by a provider: "I do get a few [patients] that tell me they're just not interested because exercise is just not for them" [P9]. Providers think that a lack of monitoring and follow-up clinical appointments could also contribute to nonadherence, with insufficient regular feedback on health progress and areas for improvement leading to decreased adherence to lifestyle changes.

Discussion

Technological Solutions

In this study, we explored the barriers to progression in the CR patient journey from the health care providers' perspectives. We identified four barriers that hinder progression in the CR patient journey, such as (1) patients not being referred to CR programs, (2) patients not enrolling in CR programs, (3) patients dropping out of CR programs, and (4) patients' lack of adherence to lifestyle changes post-CR programs. Based on our findings, we recommend 4 potential technological solutions to overcome these barriers: monitoring patients' progress throughout the CR journey using interactive data visualization systems, monitoring physical exercises through automated video analysis–based feedback systems, matching the patients to the right programs using machine learning (ML)–driven predictive analysis, and supporting patients with self-management using targeted natural language processing (NLP) and large language model (LLM) tools. These potential solutions could provide the foundation for future research and guide the development of effective technologies to overcome barriers to progression in the CR patient journey.

Monitoring Patients' Progress Throughout the CR Journey Using Interactive Data Visualization Systems

Health care providers have noticed that some patients in CR programs do not show adequate improvement in their progress, which could be due to the loss of motivation over time. In addition, they have mentioned that there may be other hidden reasons within the CR programs that are impacting patients' progress, making it difficult for patients to take necessary actions to improve progression. Effective monitoring of the patient's progress can help patients stay motivated [32-34] and help providers identify any hidden reasons affecting patients' progress [35,36]. Research has shown that interactive data visualization systems are effective in monitoring patients' progress for both patients [37,38] and providers [39,40]. Based on the success of these systems, we propose the development of interactive data visualization tools for monitoring patients' progress in CR. Patients and providers have different goals and require varying levels of detail in monitoring patients' progress [41]. Therefore, visualization systems should be developed targeting the unique needs of patients and providers with a human-centered approach [42] while involving them in the process of design and development.

For patients, visualization systems can enable them to monitor their progress throughout the CR journey by using EMR and personally collected health data in CR, such as vital signs, exercise performance, attendance, and assessment results. The system can include gamified visualizations [43] and real-life object representations [44] to make it engaging and easier for patients to understand their progress, rather than using complex charts. For example, a patient who visually sees an improvement in heart rate rhythm and blood pressure after regular exercise may feel more encouraged to actively participate throughout the CR journey.

For providers, visualization systems can assist them in monitoring the progress of patients in CR journey and help them identify any underlying reasons that affect patient progression by using aggregated EMR data of patients in the CR center. Providers can use the visualization system to filter and drill down into specific patients or groups based on demographic characteristics, particular cohorts, or specific CR sessions to analyze patients' attendance patterns and improvements or declines in progress [45]. For instance, providers might discover a trend where patients older than 65 years show better progress in heart rate rhythm when exercise sessions are coupled with meditation than when they are offered solely.

Monitoring Physical Exercises Using Automated Video Analysis–Based Feedback Systems

Providers noticed that many patients in CR struggle to perform physical exercises correctly, particularly during VCR and after completing CR programs, where direct supervision from health care providers is not available. Without this supervision and real-time feedback, patients often find it challenging to understand proper exercise techniques, leading to uncertainty about their form and an increased risk of adopting incorrect postures or unsafe practices, which could result in injuries. Recent studies show promising results for monitoring physical exercises with artificial intelligence–powered feedback systems that analyze posture [46,47] and provide real-time exercise feedback through video analysis [48-50]. Building on the success of previous artificial intelligence–powered feedback systems, we suggest developing automated video analysis–based physical exercise feedback systems to guide and monitor patients' exercise techniques and provide immediate corrective feedback.

The system can leverage pose estimation algorithms such as DensePose [51], YOLO-Pose [52], and OpenPose [53] to identify patient posture by detecting joints in real-time video. These algorithms can be trained on a dataset including correct and incorrect postures, as demonstrated and annotated by health care providers [54]. The system can then incorporate mathematical algorithms that calculate the angles between various body joints to classify whether a posture is correct or incorrect [47]. Patients can perform physical exercises while the system captures video through their web cameras. The system would analyze the video feed in real time, classifying the exercise form as correct or incorrect. If an incorrect form is

detected, the system will identify the specific area of the body that is not aligned properly and provide immediate feedback. This feedback could be delivered through text and image overlays, offering suggestions such as "You are doing a great job, but bend more," alongside visual corrections that highlight the areas needing adjustment. In addition, real-time voice prompts could be integrated to help patients quickly understand and adjust their form during the exercise. For example, if a patient is performing a lunge and their front knee extends too far forward over their toes, the system might display a red overlay on the knee area with a voice prompt saying, "Step back slightly to keep your knee aligned with your ankle."

Matching the Patients to the Right Programs Using ML-Driven Predictive Analysis

Health care providers mentioned that patients often drop out early from CR programs, which could be because they are not matched to a program that meets their needs during referral [55,56]. Providers face challenges in identifying which patients are at risk of dropping out and selecting the appropriate program that aligns with those needs, as this process is complex and time-consuming. Previous research has demonstrated the efficacy of predictive algorithms in successfully predicting dropout rates and selecting appropriate programs. For instance, predictive algorithms have been used to predict dropout rates in health care programs [57,58], such as cognitive behavioral therapy for insomnia [59], chronic disease management programs [60], rehabilitation programs [61-63], and for selecting appropriate treatment plans [64,65]. Building on previous research, we suggest that ML-driven predictive analysis algorithms can be developed to proactively identify patients at heightened risk of dropping out from CR programs based on their characteristics and historical program outcomes and suggest the most suitable CR programs for each patient.

These algorithms can be developed using various ML models. For example, logistic regression [66] or naive Bayes models [67] can assess how specific patient factors influence the likelihood of dropout. Random forests [68], or eXtreme Gradient Boosting models [69], can be used to identify which factors (or combinations of factors) are most predictive of dropouts. Gradient boosting machines [70] or support vector machines [71] can then predict dropout risk and recommend the most suitable CR programs based on the identified factors. These algorithms can be trained using historical patient data such as age, gender, location, marital status, education level, socioeconomic status, comorbidities, and previous CR program and health outcomes (eg, completion, dropout, improvement in cardiac health, readmission, or relapse). Once trained, the algorithm can be used to match new patients to the right CR program at the point of referral. By comparing the characteristics of new patients with those of previous patients who dropped out, the algorithms can pinpoint factors associated with dropout risks, such as irregular attendance or poor performance in rehabilitation sessions. The algorithm can generate risk scores that indicate the new patients' likelihood of dropout. It can then recommend the CR program that has historically yielded the best outcomes for patients with similar profiles. For instance, a 65-year-old patient with a history of diabetes and hypertension and a 90% dropout risk due to long travel distance might be matched with a program offering fewer inpatient visits or a virtual option, emphasizing dietary management and low-impact exercises based on the success of similar profiles in historical data.

Supporting Patients With Self-Management Using Targeted NLP and LLM Tools

Providers observed that many patients in CR face difficulties in self-management. This includes difficulties in independently setting their health goals and finding reliable answers to their health-related questions outside of their scheduled appointments with providers. To help patients set their health goals on their own, NLP-based automated goal-setting tools have been shown to be effective in various medical conditions such as hypertension [72], HIV [73], stroke [74], and aphasia [75]. Similarly, to help patients find answers to their health-related questions, LLM-based chatbot solutions have been shown to be successful, particularly, for patients managing chronic diseases [76], inflammatory bowel diseases [77], and breast cancer [78]. Building on the effectiveness of these tools in various medical conditions, we suggest developing targeted NLP- and LLM-based tools specifically for patients in CR. NLP-based tools, such as mobile apps or web interfaces, can assist patients in setting personal health goals in CR. LLM-integrated chatbots can help patients find answers to their questions and enhance their understanding throughout their CR journey.

For goal setting, NLP-based tools can allow patients to input detailed information about their health conditions, preferences, goals, and obstacles in their own words. Using NLP techniques, the tool can use intent recognition [79] to identify patient objectives, entity extraction [80,81] to categorize health conditions and challenges, and sentiment analysis [82,83] to assess emotional states and motivation levels. Based on this analysis, the technology can generate personalized goals using the SMART (Specific, Measurable, Achievable, Relevant, Time-bound) framework [84]. Patients can also review, refine, prioritize, and track their goals over time. For example, if a patient wants to "exercise more" but struggles with "low energy levels," the technology could suggest a goal based on the SMART framework, such as "Walk for 20 minutes, three times a week, and gradually increase the duration as energy improves."

To provide answers to patients' questions, targeted LLM-based chatbots can be trained on CR-specific educational resources, a database of frequently asked questions by patients in CR, and anonymized patient-provider conversations specific to CR. The chatbot should be capable of providing CR education, detailed explanations to patients' questions; offering empathetic responses and coping strategies for managing stress, anxiety, and other emotional challenges; and assisting patients in finding support groups and educational materials tailored to their needs. If patients have questions anytime throughout the CR patient journey, they can access the chatbot. For instance, if a patient asks how to safely continue their exercise routine at home while experiencing menstrual cramps, the chatbot could recommend gentle stretching exercises or advise reducing workout intensity while encouraging confirmation of this guidance with their provider at their next appointment [85].

Limitations

While our study aimed to capture the perspectives of health care providers from various specialties involved in CR care, the use of snowball sampling may have introduced selection bias, as participants were primarily recruited based on a research collaboration network. Although we reached thematic saturation after 7 interviews, indicating common barriers across providers from different specialties, the small sample size of 10 participants may not fully reflect the diversity of perspectives, particularly as we did not include providers from other countries. In addition, despite following rigorous methodological standards, the qualitative analysis remains interpretative by nature and may be subject to researcher bias [84]. As with any qualitative study, the rigor of our findings should be judged by their resonance and plausibility rather than their generalizability. Furthermore, our findings may not be transferable to other CR centers that differ in resources, cultures, or geographical settings [85].

Conclusions

The objective of this study was to explore the barriers to progression in the CR patient journey from the perspectives of health care providers involved in CR care. Our findings show that patients not being referred to CR programs, patients not enrolling in CR programs, patients dropping out of CR programs, and patients' lack of adherence to lifestyle changes post-CR programs are the 4 barriers to progression as perceived by providers. We also proposed 4 potential technological solutions to overcome the barriers identified through our study. Future work should focus on designing, developing, and evaluating the technological solutions proposed in this study to overcome the barriers to progression in the CR patient journey.

Acknowledgments

This study was funded by the National Research Council Aging in Place program and the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant. The authors are grateful to all the participants for sharing their invaluable perspectives, without whom this study would not have been possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (COnsolidated criteria for REporting Qualitative research) checklist.
[PDF File (Adobe PDF File), 74 KB - ijmr_v14i1e66164_app1.pdf]

Multimedia Appendix 2
Semistructured interview script and questions.
[PDF File (Adobe PDF File), 73 KB - ijmr_v14i1e66164_app2.pdf]

Multimedia Appendix 3
Representative quotes on barriers to progression in the cardiac rehabilitation patient journey.
[PDF File (Adobe PDF File), 128 KB - ijmr_v14i1e66164_app3.pdf]

Multimedia Appendix 4
Saturation table.
[PDF File (Adobe PDF File), 135 KB - ijmr_v14i1e66164_app4.pdf]

Multimedia Appendix 5
Codebook.
[PDF File (Adobe PDF File), 81 KB - ijmr_v14i1e66164_app5.pdf]

References

1. Cardiovascular diseases (CVDs). World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2024-08-15]
2. Taylor RS, Dalal HM, McDonagh STJ. The role of cardiac rehabilitation in improving cardiovascular outcomes. Nat Rev Cardiol 2022;19(3):180-194 [FREE Full text] [doi: 10.1038/s41569-021-00611-7] [Medline: 34531576]
3. Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, American Heart Association Exercise, Cardiac Rehabilitation,Prevention Committee, the Council on Clinical Cardiology, American Heart Association Council on Cardiovascular Nursing, American Heart Association Council on EpidemiologyPrevention, American Heart Association Council on Nutrition, Physical Activity,Metabolism, American Association of CardiovascularPulmonary Rehabilitation. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the

American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007;115(20):2675-2682. [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
4. Protoview P, Vol B, Ringgold B. Guidelines for Cardiac Rehabilitation Programs, 6th Edition. Champaign, IL: Human Kinetics; 2020.
5. Long L, Mordi I, Bridges C, Sagar VA, Davies EJ, Coats AJS, et al. Exercise-based cardiac rehabilitation for adults with heart failure. Cochrane Database Syst Rev 2019;1(1):CD003331 [FREE Full text] [doi: 10.1002/14651858.CD003331.pub5] [Medline: 30695817]
6. Anderson L, Thompson DR, Oldridge N, Zwisler A, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016;2016(1):CD001800 [FREE Full text] [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]
7. Jiménez-Navarro MF, Lopez-Jimenez F, Pérez-Belmonte LM, Lennon RJ, Diaz-Melean C, Rodriguez-Escudero JP, et al. Benefits of cardiac rehabilitation on cardiovascular outcomes in patients with diabetes mellitus after percutaneous coronary intervention. J Am Heart Assoc 2017;6(10):e006404 [FREE Full text] [doi: 10.1161/JAHA.117.006404] [Medline: 29021270]
8. Yohannes AM, Doherty P, Bundy C, Yalfani A. The long-term benefits of cardiac rehabilitation on depression, anxiety, physical activity and quality of life. J Clin Nurs 2010;19(19-20):2806-2813. [doi: 10.1111/j.1365-2702.2010.03313.x] [Medline: 20738450]
9. Campbell M, McEvoy JW, Calpin G, Concannon F, Redfern S. Perspectives of healthcare professionals on clinician-patient communication of cardiovascular disease risk. J Patient Exp 2024;11:23743735241257386 [FREE Full text] [doi: 10.1177/23743735241257386] [Medline: 38807917]
10. Vanzella LM, Cotie LM, Flores-Hukom M, Marzolini S, Konidis R, Ghisi GLDM. Patients' perceptions of hybrid and virtual-only care models during the cardiac rehabilitation patient journey: a qualitative study. J Cardiovasc Nurs 2025;40(2):E91-E100. [doi: 10.1097/JCN.0000000000001076] [Medline: 38206327]
11. Jackson L, Leclerc J, Erskine Y, Linden W. Getting the most out of cardiac rehabilitation: a review of referral and adherence predictors. Heart 2005;91(1):10-14 [FREE Full text] [doi: 10.1136/hrt.2004.045559] [Medline: 15604322]
12. Ritchey MD, Maresh S, McNeely J, Shaffer T, Jackson SL, Keteyian SJ, et al. Tracking cardiac rehabilitation participation and completion among medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020;13(1):e005902 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
13. Janssen V, De Gucht V, van Exel H, Maes S. A self-regulation lifestyle program for post-cardiac rehabilitation patients has long-term effects on exercise adherence. J Behav Med 2014;37(2):308-321. [doi: 10.1007/s10865-012-9489-y] [Medline: 23334387]
14. Platz K, Kools S, Howie-Esquivel J. Benefits, facilitators, and barriers of alternative models of cardiac rehabilitation: a qualitative systematic review. J Cardiopulm Rehabil Prev 2023;43(2):83-92. [doi: 10.1097/HCR.0000000000000738] [Medline: 36346781]
15. Sérvio TC, Britto RR, de Melo Ghisi GL, da Silva LP, Silva LDN, Lima MMO, et al. Barriers to cardiac rehabilitation delivery in a low-resource setting from the perspective of healthcare administrators, rehabilitation providers, and cardiac patients. BMC Health Serv Res 2019;19(1):615 [FREE Full text] [doi: 10.1186/s12913-019-4463-9] [Medline: 31477103]
16. Resurrección DM, Motrico E, Rigabert A, Rubio-Valera M, Conejo-Cerón S, Pastor L, et al. Barriers for nonparticipation and dropout of women in cardiac rehabilitation programs: a systematic review. J Womens Health (Larchmt) 2017;26(8):849-859. [doi: 10.1089/jwh.2016.6249] [Medline: 28388314]
17. Ozemek C, Squires RW. Enrollment and adherence to early outpatient and maintenance cardiac rehabilitation programs. J Cardiopulm Rehabil Prev 2021;41(6):367-374. [doi: 10.1097/HCR.0000000000000645] [Medline: 34727555]
18. Supervía M, Medina-Inojosa JR, Yeung C, Lopez-Jimenez F, Squires RW, Pérez-Terzic CM, et al. Cardiac rehabilitation for women: a systematic review of barriers and solutions. Mayo Clin Proc 2017;92(4):565-577 [FREE Full text] [doi: 10.1016/j.mayocp.2017.01.002] [Medline: 28365100]
19. Gravely-Witte S, Leung YW, Nariani R, Tamim H, Oh P, Chan VM, et al. Effects of cardiac rehabilitation referral strategies on referral and enrollment rates. Nat Rev Cardiol 2010;7(2):87-96 [FREE Full text] [doi: 10.1038/nrcardio.2009.223] [Medline: 19997077]
20. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the million hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017;92(2):234-242 [FREE Full text] [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
21. O'Doherty AF, Humphreys H, Dawkes S, Cowie A, Hinton S, Brubaker PH, et al. How has technology been used to deliver cardiac rehabilitation during the COVID-19 pandemic? An international cross-sectional survey of healthcare professionals conducted by the BACPR. BMJ Open 2021;11(4):e046051 [FREE Full text] [doi: 10.1136/bmjopen-2020-046051] [Medline: 33879492]
22. Beatty AL, Beckie TM, Dodson J, Goldstein CM, Hughes JW, Kraus WE, et al. A new era in cardiac rehabilitation delivery: research gaps, questions, strategies, and priorities. Circulation 2023;147(3):254-266 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.122.061046] [Medline: 36649394]

23. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
24. Britten N. Qualitative interviews in medical research. BMJ 1995;311(6999):251-253 [FREE Full text] [doi: 10.1136/bmj.311.6999.251] [Medline: 7627048]
25. Patton MQ. Qualitative Evaluation and Research Methods, 2nd ed. Thousand Oaks, CA: Sage Publications, Inc; 1990.
26. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. Soc Methods Res 1981;10(2):141-163 [FREE Full text] [doi: 10.1177/004912418101000205]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
28. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995;311(6996):42-45 [FREE Full text] [doi: 10.1136/bmj.311.6996.42] [Medline: 7613329]
29. Ahmed SK. The pillars of trustworthiness in qualitative research. J. Public Health Med 2024;2:100051. [doi: 10.1016/j.glmedi.2024.100051]
30. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Dir Program Eval 2004;1986(30):73-84. [doi: 10.1002/ev.1427]
31. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
32. Cajamarca G, Herskovic V, Rossel PO. Enabling older adults' health self-management through self-report and visualization—a systematic literature review. Sensors (Basel) 2020;20(15):4348 [FREE Full text] [doi: 10.3390/s20154348] [Medline: 32759801]
33. Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, et al. Data visualization for chronic neurological and mental health condition self-management: systematic review of user perspectives. JMIR Ment Health 2022;9(4):e25249 [FREE Full text] [doi: 10.2196/25249] [Medline: 35482368]
34. Maitland J, Chalmers M. Self-monitoring, self-awareness, and self-determination in cardiac rehabilitation. 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2010 April 10; Atlanta, GA p. 1213-1222. [doi: 10.1038/s41569-021-00611-7]
35. McCarthy S, O'Raghallaigh P, Woodworth S, Lim YL, Kenny LC, Adam F. An integrated patient journey mapping tool for embedding quality in healthcare service reform. J Decis Syst 2016;25(sup1):354-368. [doi: 10.1145/1753326.1753508]
36. Borycki EM, Kushniruk AW, Wagner E, Kletke R. Patient journey mapping: integrating digital technologies into the journey. Knowl Manag e-learn Int J 2020;12(4):521-535. [doi: 10.34105/j.kmel.2020.12.029]
37. Chen L, Yan N, Kiang M, Muth A, Krishna K. Innomotion: a web-based rehabilitation system helping patients recover and gain self-awareness of their body away from the clinic. 2014 Presented at: CHI '14 Extended Abstracts on Human Factors in Computing Systems; 2014 April 26; Ontario, Toronto, Canada p. 233-238. [doi: 10.1145/2559206.2580926]
38. Zhu J, Lei Y, Shah A, Schein G, Ghaednia H, Schwab J, et al. MuscleRehab: Improving unsupervised physical rehabilitation by monitoring and visualizing muscle engagement. 2022 Presented at: UIST '22: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology; 2022 October 28; Bend, OR p. 1-14. [doi: 10.1145/3526113.3545705]
39. Shafei I, Karnon J, Crotty M. Process mining and customer journey mapping in healthcare: enhancing patient-centred care in stroke rehabilitation. Digit Health 2024;10:20552076241249264 [FREE Full text] [doi: 10.1177/20552076241249264] [Medline: 38766357]
40. Widanagamaachchi W, Livnat Y, Bremer P, Duvall S, Pascucci V. Interactive visualization and exploration of patient progression in a hospital setting. AMIA Annu Symp Proc 2017;2017:1773-1782 [FREE Full text] [Medline: 29854248]
41. Rajabiyazdi F, Perin C, Oehlberg L, Carpendale S. Exploring the design of patient-generated data visualizations. 2020. URL: https://hal.science/hal-02861239 [accessed 2024-08-11]
42. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
43. Feng Z, Gao Y, Zhang T. Gamification for visualization applications in the construction industry. Struct Integr 2022;20:495-514. [doi: 10.1007/978-3-030-82430-3_21]
44. Urbano ICVP, Guerreiro JPV, Nicolau HMAA. From skeuomorphism to flat design: age-related differences in performance and aesthetic perceptions. Behav Inform Technol 2020;41(3):452-467. [doi: 10.1080/0144929X.2020.1814867]
45. Loorak MH, Perin C, Kamal N, Hill M, Carpendale S. TimeSpan: using visualization to explore temporal multi-dimensional data of stroke patients. IEEE Trans Vis Comput Graph 2016;22(1):409-418 [FREE Full text] [doi: 10.1109/TVCG.2015.2467325] [Medline: 26390482]
46. Sowmia KR, Jayaganeshan T, Khan MA, Madhesh S, Kabilesh S. An artificial intelligence approach to quantifying exercise form for optimal performance and injury prevention. : Springer; 2024 Presented at: Proceedings of Third International Conference on Computing and Communication Networks Lecture Notes in Networks and Systems; 2024; Singapore p. 639-647. [doi: 10.1007/978-981-97-0892-5_50]

47. Ramesh SH, Lemaire ED, Tu A, Cheung K, Baddour N. Automated implementation of the Edinburgh Visual Gait Score (EVGS) using openpose and handheld smartphone video. Sensors (Basel) 2023;23(10):4839 [FREE Full text] [doi: 10.3390/s23104839] [Medline: 37430751]
48. Kotte H, Daiber F, Kravcik M, Duong-Trung N. FitSight: tracking and feedback engine for personalized fitness training. 2024 Presented at: UMAP '24: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization; 2024 June 22; Cagliari, Italy p. 223-231. [doi: 10.1145/3627043.3659547]
49. Postolache O, Hemanth DJ, Alexandre R, Gupta D, Geman O, Khanna A. Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality. IEEE J Select Areas Commun 2021;39(2):562-573. [doi: 10.1109/jsac.2020.3020600]
50. Chatzitofis A, Zarpalas D, Filos D, Triantafyllidis A, Chouvarda I, Maglaveras N, et al. Technological module for unsupervised, personalized cardiac rehabilitation exercising. 2017 Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017 July 04-08; Turin, Italy p. 125-130. [doi: 10.1109/COMPSAC.2017.230]
51. Güler RA, Neverova N, Kokkinos I. DensePose: dense human pose estimation in the wild. 2018 Presented at: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2018 June 18-23; Salt Lake City, UT p. 7297-7306. [doi: 10.1109/CVPR.2018.00762]
52. Maji D, Nagori S, Mathew M, Poddar D. YOLO-Pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss. 2022 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2022 June 19-20; New Orleans, LA p. 2636-2645. [doi: 10.1109/CVPRW56347.2022.00297]
53. Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 2021;43(1):172-186. [doi: 10.1109/TPAMI.2019.2929257] [Medline: 31331883]
54. Rangari T, Kumar S, Roy PP, Dogra DP, Kim B. Video based exercise recognition and correct pose detection. Multimed Tools Appl 2022;81(21):30267-30282. [doi: 10.1007/s11042-022-12299-z]
55. Elsakr C, Bulger DA, Roman S, Kirolos I, Khouzam RN. Barriers physicians face when referring patients to cardiac rehabilitation: a narrative review. Ann Transl Med 2019;7(17):414 [FREE Full text] [doi: 10.21037/atm.2019.07.61] [Medline: 31660313]
56. Ghisi GLM, Polyzotis P, Oh P, Pakosh M, Grace SL. Physician factors affecting cardiac rehabilitation referral and patient enrollment: a systematic review. Clin Cardiol 2013;36(6):323-335 [FREE Full text] [doi: 10.1002/clc.22126] [Medline: 23640785]
57. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. J Med Internet Res 2019;21(9):e13617 [FREE Full text] [doi: 10.2196/13617] [Medline: 31486409]
58. Bremer V, Chow PI, Funk B, Thorndike FP, Ritterband LM. Developing a process for the analysis of user journeys and the prediction of dropout in digital health interventions: machine learning approach. J Med Internet Res 2020;22(10):e17738 [FREE Full text] [doi: 10.2196/17738] [Medline: 33112241]
59. Bennemann B, Schwartz B, Giesemann J, Lutz W. Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. Br J Psychiatry 2022;220(4):192-201. [doi: 10.1192/bjp.2022.17] [Medline: 35177132]
60. Lee S, Chung JY. The machine learning-based dropout early warning system for improving the performance of dropout prediction. Appl Sci 2019;9(15):3093. [doi: 10.3390/app9153093]
61. Gross DP, Zhang J, Steenstra I, Barnsley S, Haws C, Amell T, et al. Development of a computer-based clinical decision support tool for selecting appropriate rehabilitation interventions for injured workers. J Occup Rehabil 2013;23(4):597-609. [doi: 10.1007/s10926-013-9430-4] [Medline: 23468410]
62. Lofaro D, Groccia M, Guido R, Conforti D, Caroleo S, Fragomeni G. Machine learning approaches for supporting patient-specific cardiac rehabilitation programs. Comput Cardiol (2010) 2016;43:149-152. [doi: 10.22489/CinC.2016.047-209]
63. Nithya B, Ilango V. Predictive analytics in health care using machine learning tools and techniques. 2017 Presented at: Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS; 2017 June 15-16; Madurai, India p. 492-499. [doi: 10.1109/ICCONS.2017.8250771]
64. Valdes G, Simone CB, Chen J, Lin A, Yom SS, Pattison AJ, et al. Clinical decision support of radiotherapy treatment planning: a data-driven machine learning strategy for patient-specific dosimetric decision making. Radiother Oncol 2017;125(3):392-397. [doi: 10.1016/j.radonc.2017.10.014] [Medline: 29162279]
65. Swarnakar R, Yadav SL. Artificial intelligence and machine learning in motor recovery: a rehabilitation medicine perspective. World J Clin Cases 2023;11(29):7258-7260 [FREE Full text] [doi: 10.12998/wjcc.v11.i29.7258] [Medline: 37946764]
66. Fahrmeir L, Kneib T, Lang S, Marx B. Regression models. In: Regression. Berlin, Heidelberg: Springer; 2013:21-72. [doi: 10.1007/978-3-642-34333-9]
67. Yang FJ. An implementation of naive bayes classifier. 2018 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); 2018 December 12-14; Las Vegas, NV p. 301-306. [doi: 10.1109/CSCI46756.2018.00065]
68. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]

69. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13; San Francisco, CA p. 785-794. [doi: https://doi.org/10.1145/2939672.2939785]
70. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21 [FREE Full text] [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]
71. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl 1998;13(4):18-28. [doi: 10.1109/5254.708428]
72. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
73. Mozafarinia M, Rajabiyazdi F, Brouillette M, Fellows LK, Mayo NE. Development and usability of a feedback tool, "My Personal Brain Health Dashboard", to improve setting of self-management goals among people living with HIV in Canada. Qual Life Res 2021;30(11):3199-3211. [doi: 10.1007/s11136-020-02555-w] [Medline: 32556825]
74. Ha Y, Park M. Effects of stroke rehabilitation using gait robot-assisted training and person-centered goal setting: a single blinded pilot study. Healthcare (Basel) 2023;11(4):588 [FREE Full text] [doi: 10.3390/healthcare11040588] [Medline: 36833122]
75. Hersh D, Worrall L, Howe T, Sherratt S, Davidson B. SMARTER goal setting in aphasia rehabilitation. Aphasiology 2012;26(2):220-233. [doi: 10.1080/02687038.2011.640392]
76. Chaix B, Guillemassé A, Nectoux P, Delamon G, Brouard B. Vik: a chatbot to support patients with chronic diseases. Health N Hav 2020;12(07):804-810. [doi: 10.4236/health.2020.127058]
77. Zand A, Sharma A, Stokes Z, Reynolds C, Montilla A, Sauk J, et al. An exploration into the use of a chatbot for patients with inflammatory bowel diseases: retrospective cohort study. J Med Internet Res 2020;22(5):e15589 [FREE Full text] [doi: 10.2196/15589] [Medline: 32452808]
78. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
79. Khan V, Meenai TA. Pretrained natural language processing model for intent recognition (BERT-IR). Hum Centric Intell Syst 2021;1(3):66-74. [doi: 10.2991/hcis.k.211109.001]
80. Finkel J, Grenager T, Manning C. Incorporating non-local information into information extraction systems by Gibbs sampling. 2005 Presented at: ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics; June 25, 2005; Ann Arbor, MI p. 363-370. [doi: 10.3115/1219840.1219885]
81. Akbik A, Blythe D, Vollgraf R. Contextual string embeddings for sequence labeling. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics; 2018 August 20-26; Santa Fe, NM p. 1638-1649 URL: https://aclanthology.org/C18-1139/
82. Glazkova A, Glazkov M, Trifonov T. g2tmn at Constraint@AAAI2021xploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. Commun Comput Inf Sci 2021:116-127. [doi: 10.1007/978-3-030-73696-5_12]
83. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. 2018 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 01, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.1038/s41569-021-00611-7]
84. Doran G. There's a S.M.A.R.T. way to write managements's goals and objectives. Manage Rev 1981;70(11):35-36 [FREE Full text]
85. Anandan P, Kokila S, Elango S, Gopinath P, Sudarsan P. Artificial intelligence based chat bot for patient health care. 2022 Presented at: International Conference on Computer Communication and Informatics (ICCCI); 2022 January 25-27; Coimbatore, India. [doi: 10.1109/ICCCI54379.2022.9740912]

Abbreviations

COREQ: COnsolidated criteria for REporting Qualitative research
CR: cardiac rehabilitation
EMR: electronic medical record
LLM: large language model
ML: machine learning
NLP: natural language processing
SMART: Specific, Measurable, Achievable, Relevant, Time-bound
VCR: virtual cardiac rehabilitation

Edited by T de Azevedo Cardoso; submitted 05.09.24; peer-reviewed by M Mohammadnezhad, J Walsh; comments to author 04.12.24; revised version received 15.01.25; accepted 21.01.25; published 21.02.25.

Please cite as:
Ramesh SH, Jull D, Fournier H, Rajabiyazdi F
Exploring Barriers to Patients' Progression in the Cardiac Rehabilitation Journey From Health Care Providers' Perspectives: Qualitative Study
Interact J Med Res 2025;14:e66164
URL: https://www.i-jmr.org/2025/1/e66164
doi:10.2196/66164
PMID:

Long-Term Engagement of Diverse Study Cohorts in Decentralized Research: Longitudinal Analysis of "All of Us" Research Program Data

Vijay Yadav, MS; Elias Chaibub Neto, PhD; Megan Doerr, MS; Abhishek Pratap, PhD; Larsson Omberg, PhD

Sage Bionetworks, 2901 3rd Ave #330, Seattle, WA, United States

Corresponding Author:
Vijay Yadav, MS
Sage Bionetworks, 2901 3rd Ave #330, Seattle, WA, United States

Abstract

Background: The generalizability of clinical research hinges on robust study designs, which include the recruitment and maintenance of a representative study population. This study examines the evolution of the demographic characteristics of 329,038 participants who enrolled and participated in The *All of Us* Research Program (AoURP), a decentralized study aimed at representing the diversity of the United States.

Objective: The primary objectives of this study were to assess alterations in the demographic composition of the cohort at different protocol stages within AoURP, while analyzing completion rates and timeframes for survey and substudy completion. Additionally, we examined how participant interactions with the program impacted engagement and survey responses.

Methods: We conducted a longitudinal analysis of the AoURP data, tracking changes in demographic composition, completion rates, and completion times for surveys and substudies. Comparative analyses were performed to assess differences in engagement and survey completion based on sociodemographic characteristics of participants involved in postenrollment study components.

Results: The sociodemographic composition of the cohort that participated in the postenrollment study (eg, optional components) differed significantly from that of the recruited population. The proportion of self-identified White participants increased by 21.2%, whereas the proportion of Black or African American participants decreased by 12.18% (P=.02). Participants who identified as White (n=93,614, 52.7%) and NonHispanic (n=109,279, 42.21%) were more engaged compared to those identifying as Black or African American (n=10,887, 15.76%), Asian (n=4274, 38.72%), or Hispanic (n=12,530, 20.7%; P=.006). Participants' response times to study surveys and completeness varied across all demographic groups (P<.001). Furthermore, those identifying as White skipped fewer survey questions (1.19) compared to those identifying as Black or African American (1.40) or other racial and ethnic identities (P<.001).

Conclusions: The AoURP dataset serves as an exceptional resource for investigating diverse public health concerns. However, the longitudinal analysis of participant-level data underscores a significant skew in population diversity, suggesting the need for targeted strategies to enhance engagement and retention across all groups. Ensuring diversity in the cohort is essential for maintaining the study's representativeness and the broad applicability of its findings.

(Interact J Med Res 2025;14:e56803) doi:10.2196/56803

KEYWORDS

digital health; engagement behavior; All of Us Research Program; retention; decentralized research cohorts

Introduction

There is a history of lack of racial, ethnic, and gender diversity in health studies [1,2]. This lack of diversity can lead to issues related to the generalizability of research findings and equity in health care. However, lack of plurality has decreased in the last three decades, partially due to regulatory and policy efforts within government agencies, such as the National Institutes of Health [3], Food and Drug Administration [4], and Department of Health Services [5], that have sought to enhance minority participation in clinical research as well as identified scientific need [6]. As the demographic makeup of the United States is becoming more pluralistic, diverse study populations are necessary to have representative and translatable observations. Much of the effort to increase minority representation in research has focused on the recruitment of diverse study populations. However, the recruited population is not always representative of the population that is ultimately studied, due to engagement rates being different across sociodemographic groups. This has been demonstrated in remote observational studies, where there is less direct interaction with participants compared to studies which require in-person visits [7]. Poor engagement can lead to study failure due to reduced sample sizes, causing loss of power or imbalanced study populations. Prior research has

shown that remote-only studies demonstrate different engagement rates based on demographic features including disease presence or absence, age, race or ethnicity, and recruitment methods [6-8]. When the engagement rates vary across different populations, there is a risk that the effective study population (ie, the population with longitudinal data) becomes unrepresentative of the originally recruited cohort.

The *All of Us* Research Program (AoURP) was launched in May 2018, with an aim to recruit more than 1 million participants living in the United States to accelerate health research and precision medicine. AoURP has specifically focused on recruiting demographic categories that have historically been underrepresented in biomedical research (UBRs) and has largely succeeded in this objective through partnership with more than 340 recruitment sites nationwide [3]. During the first year of the AoURP, 80% of recruited participants self-identified as belonging to one or more UBR populations [9]. In this study, we explore long-term engagement within AoURP by exploring participation in optional components of the AoURP study (eg, surveys and substudies that could be performed post enrollment), the time that it takes participants to complete optional surveys, and survey response completeness. We hypothesize that this information, combined with data on specific interactions of participants with the AoURP can be used to improve and develop strategies that promote sustained engagement across the diverse demographics. This study primarily assesses changes in cohort demographics across AoURP protocol stages, completion rates for surveys and substudies, and the influence of participant interactions on engagement and responses. These insights provide actionable strategies for sustaining diverse, representative cohorts.

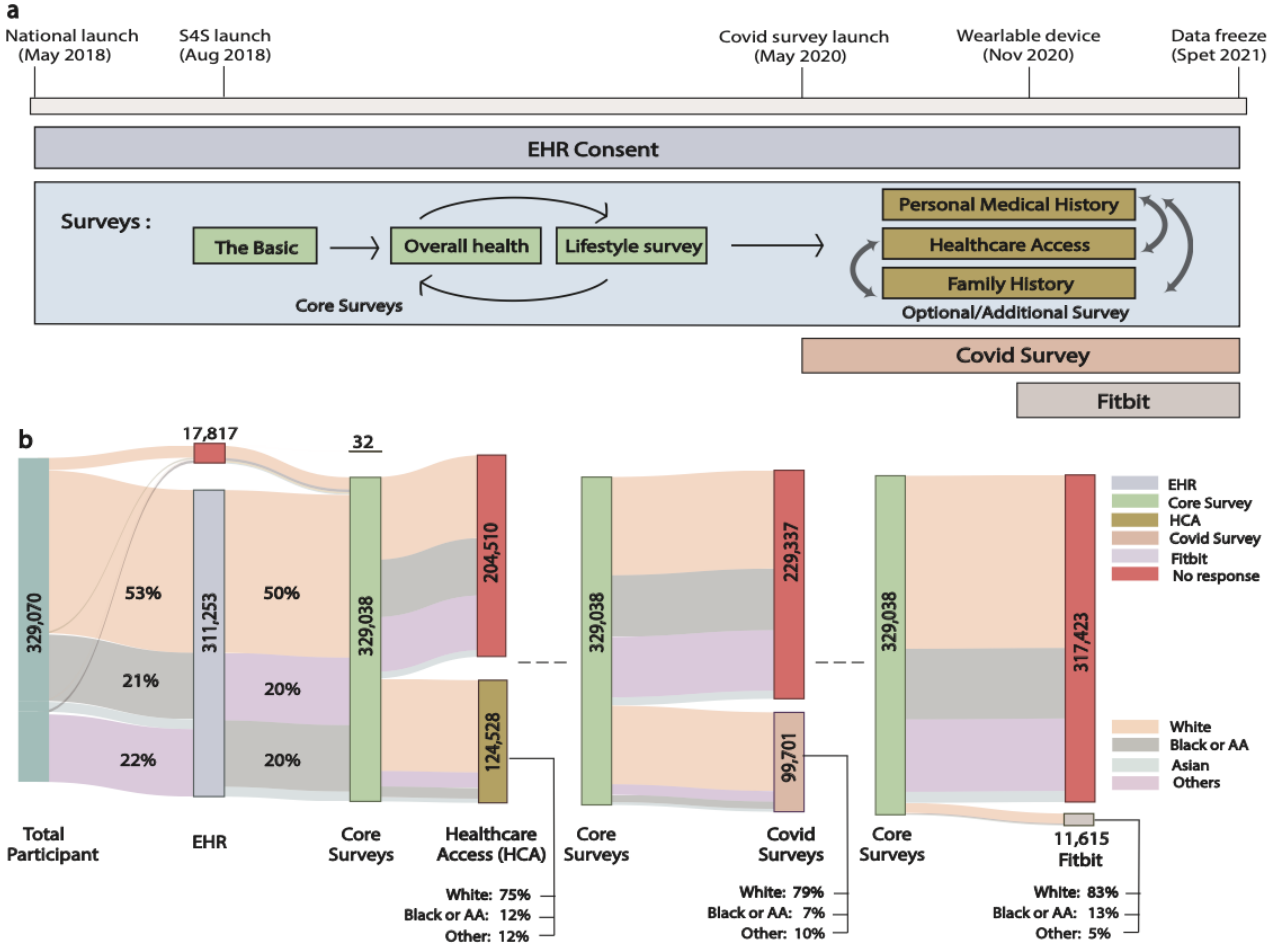
Methods

Study Design and Participants

This study is a longitudinal study that analyzes data from participants in the AoURP, encompassing individuals aged 18 and older residing in the United States, irrespective of race, ethnicity, sex, gender, or sexual orientation [10,11]. Data from the 329,038 participants in the September 2021 AoURP data freeze were analyzed. The AoURP protocol has evolved since its launch in 2018 (Figure 1A). At present, after consenting, participants are given the option to share their electronic health record (EHR) data, provide a biosample, and answer 3 core surveys (ie, The Basics, Overall Health, and Lifestyle) [12]. Participants can also respond to additional optional health surveys including Healthcare Access (HCA), Personal Medical History (PMH), and Family Medical History (FMH), at any time and in any order after enrollment (Table S1 in Multimedia Appendix 1) [12]. In addition to these components, AoURP included two additional optional substudies, at the time of analysis: (1) a COVID-19 survey and (2) a "bring your own device" (BYOD) Fitbit study, which allows participants to share data from any Fitbit wearable device (Google Inc, Mountain View, CA) owned by them. The AoURP also collects additional data, such as physical measurements; however, our analysis is restricted to recruitment and engagement characteristics based on participant-provided information (eg, health and demographic surveys) and BYOD Fitbit data.

The core surveys included information about the participants' self-reported demographics, which was used to explore differences in participation. Specifically, we used responses to the questions on (1) date of birth; (2) What was your biological sex assigned at birth; (3) Which categories describe you (race); (4) Which categories describe you (ethnicity); and (5) What is your annual household income (from all sources) [12]. Based on the *All of Us* (AoU) Researcher Workbench, we categorized age into three bands: younger (18 - 44 years), middle aged (45 - 64 years), and older adults (≥65), and household income into 4 bands (<$50,000, $50,000-$100,000, $100,000-$200,000, >$200,000). For other multiple-choice questions, we included categories with at least 1.8% respondents. This resulted in the following categories: male and female for biological sex at birth; White, Black or African American, and Asian for race; and NonHispanic and Hispanic for ethnicity. These categories were used to assess changes in cohort demographics during the course of the study, with the primary outcomes under investigation being (1) engagement (ie, number of participants who remain engaged by completing optional components after enrollment), (2) response time (ie, time between invitation and completion of optional components), and (3) completeness of response (ie, how often participants chose to answer specific questions). Understanding these changes is essential for examining shifts in participant demographics and engagement over time, which helps identify trends and biases, ensuring the cohort representativeness and development of targeted strategies for improved engagement and retention.

Figure 1. Protocol and changes in engagement across components in study. (A) Timeline for components of protocol and journey of a participant in the AoU program. Based on respective protocol launch dates, participants had the option to consent to share EHR data, participate in COVID-19 surveys, and share data from a BYOD Fitbit data. Each survey block reflects the ordering of core and additional surveys configured in the program. (B) Sankey diagram showing the population distribution within the AoURP throughout the protocol journey. The nodes represent different types of events (eg, EHR consent, completion of optional surveys, BYOD Fitbit data sharing). Flows between notes are color-coded by self-identified race (ie, White, Black, or African American, Asian, or other). The width of the nodes and links provide quantitative information. AA: African American, AoURP: *All of Us* Research Program; BYOD: bring your own device; EHR: electronic health records; HCA: Healthcare Access survey.



Data Processing

For the primary analysis, we built an engagement cohort consisting of participants who successfully enrolled in the study and responded to the demographic questions using the AoU Researcher Workbench [13]. Since all components of the study were not launched at the same time (eg, BYOD Fitbit and the sharing of electronic health records using Sync4Science [14] were added later; see Figure 1A for the timeline), we conducted a secondary analysis using enrollment dates extracted from the observation table to evaluate the effect of timing of enrollment on these components. Specifically, this secondary analysis included individuals who enrolled after Sync4Science launch (August, 2018) and after the BYOD Fitbit protocol being launched (November, 2020). For all cohorts, activities where participants completed the surveys were considered as "active/engaged" events. All survey questions included a "prefer not to answer" option. Survey completeness was analyzed across participants by counting the number of questions answered as "prefer not to answer".

Statistical Analysis

Various statistical techniques were used for the three different analyses. Whenever possible we favored the use of nonparametric statistical tests and resampling techniques to avoid relying on distributional assumptions. These included the Kruskal-Wallis test (instead of analysis of variance), Mann-Whitney tests for pairwise group comparisons), bootstrapping for computation of confidence intervals, and permutation tests for calculating statistical significance. For all analyses, a significance level of $P<.05$ was used with Bonferroni correction for multiple testing [15].

To analyze engagement across self-reported age, sex at birth, race, household income, and ethnicity groups, we compared the proportion of participants responding in each demographic group between the core survey and optional components using a χ^2 test for homogeneity.

Participants' response time, as measured by the number of days between joining the study and completing (or joining) an optional component, was analyzed using the nonparametric Kruskal-Wallis one-way analysis of variance for each optional

component separately, as the data was not normally distributed [16]. To further understand the variability between each group, pairwise differences between groups were analyzed using the Python language (version number 3.9; Python Software Foundation), including the *statannotation* package (version 0.4.2) [17], along with the with Mann-Whitney integrated statistical test and Bonferroni for multiple testing correction [18,19].

To better understand the relationship between various self-reported demographic variables, a linear mixed model (LMM) was used to analyze the response time across the 3 surveys concomitantly. To improve the fit of the model, response time data was log-transformed before feeding into the model. All self-reported demographic variables were fitted into the model to evaluate the effect produced by each predictor. Age was treated as a continuous variable, while other variables were fitted as categorical variables. The coefficient of relative change for each variable was reported by the LMM model, and the log-transformed relative change between each group was back-transformed to a percentage change in response time across groups for all the categories, using formula $100(e^{\text{coefficient}}-1)$. To further evaluate confidence intervals, bootstrapping was used. To account for the repeated measurements of each subject, the bootstrap confidence intervals for the LMMs were constructed based on 1000 iterations of the following couple of steps. First, we obtained a bootstrap sample from the original data by sampling with replacement at the subject level (ie, assuming that our dataset contains data from *N* subjects, we grouped the repeated measurements of each subject as a "subject block," and then sampled with replacement N "subject blocks" out of the N blocks in the original dataset.) Second, for each bootstrap sample, we fit a LMM and computed the coefficient of relative change of each variable as described above. Finally, we computed 95% CI for the coefficients of relative change using the percentile interval method, where the lower and upper confidence bounds correspond to the 0.025 and 0.975 quantiles of the distribution of coefficients generated across the 1000 bootstrap iterations.

We independently examined outliers in the response time for optional surveys to check if these outliers were the cause of any systematic differences. Outliers comprised any individual who had a response time less than the first quartile (Q1)–1.5×IQR or greater than Q3+1.5×IQR. To evaluate the differences, we computed the proportion of outlier participants for each demographic group and then conducted a χ^2 test (test of homogeneity) for testing the equality of outlier participants proportion across different demographic groups.

Completeness of response was evaluated by counting the number of questions where a participant responded "prefer not to answer." First, the nonparametric Kruskal-Wallis test was used to evaluate variability across different groups (since the data was not normally distributed). However, due to ties in the completeness of the response variable, we adopted a permutation test to assess differences between the groups [12]. The permutation test was conducted by comparing the observed value of the ANOVA F-statistic computed in the original data (F_{obs}) against the permutation null distribution of the F-statistic computed on permuted versions of the data (F*), where the group labels were randomly shuffled. The test was based on 1000 data permutations and the permutation *P* value was computed as the proportion of times the permuted F-statistic was greater or equal to the observed test statistic (ie, $P=(1+\text{sum}\{i=1..B \text{ indicator}\{F^*>=F_{obs}\}\})/(1+B)$, for B=1000). Due to the large sample set, it is possible that we may be detecting very small effect sizes with very high statistical significance across different categories. To assess the effect size, 95% CI were computed using nonparametric bootstrapping for the pairwise group mean difference for each group [20].

Ethical Considerations

This study used deidentified data from the *All of Us* Research Program, which obtained institutional review board (IRB) approval and informed consent from participants for secondary research [21]. The original data collection was approved by the AoURP IRB [21], and no additional IRB approval was required for this secondary analysis. According to the All of Us Responsible Conduct of Research training, analyses using deidentified data on the Researcher Workbench do not require a separate IRB review [22], as the research involves no direct interaction with participants. All analyses adhered to ethical principles, ensuring privacy and confidentiality as outlined by the AoURP.

Results

Population Characteristics

Data from 329,078 consented participants were made available as part of the version 5 data release [23] by AoURP. Of the total consented participants, 329,038 (99.98%) completed all core surveys and were considered successfully enrolled in the program. Additionally, 311,253 participants (94.59%) consented to share all their electronic health records; including retrospective medical records from before the AoURP study was launched. Post completion of core surveys, participants had the option to complete three optional health surveys. A significantly smaller proportion completed these three optional surveys: (1) HCA-37.84% (n=124,528/329,078), (2) FMH-117,693 (35.76%), and (3) PMH-113,830 (34.59%) (Figure 1B). Both the COVID survey and the BYOD Fitbit data collection were launched at later stages of the study (Figure 1A). A total of 99,701 (30.3%) people participated in the first COVID survey and 11,615 (3.52%) participated in BYOD Fitbit data sharing. Participants who enrolled before the BYOD Fitbit data collection protocol was launched were less likely to consent to share their BYOD Fitbit data compared to those who enrolled after the protocol launch (3.51% vs 6.51%, respectively).

The AoURP aims to recruit persons from the UBR populations, and recruitment was successful, with 24.36% (n=80,154) of the participants responding to the core surveys with a self-reported race of Black or African American and Asian, while 53.98% (n=177,615) self-reported as White. We compared the distribution of ages and race to the 2020 US Census data [24], which showed the enrichment of self-identified races other than White, as well as middle-aged individuals, in the AoU cohort (Figure 2A) at baseline. Additionally, more participants reported

being female (Figure S1a in Multimedia Appendix 1) and Hispanic ethnicity (Figure S1b in Multimedia Appendix 1) compared to the census.(However, this initial enrichment of people self-identifying as belonging to one or more UBR populations diminished as participants engaged with optional components of the protocol. Distributions across demographic variables for each component of the study is shown in Table 1. Moreover, a total of 1416 (0.4%) deaths were reported by the end of version 5 data release.

Figure 2. (a) Population pyramids divided by self-reported race (White vs other) see Multimedia Appendix 1 for additional populations pyramids. Three different pyramids are displayed for comparison for three different populations representing: those who responded to: AoURP core survey, one optional survey (HCA) and the 2020 US census. At baseline, based on response to the core survey) cohorts tend to be older and less white compared to the 2020 US Census, while responses to later optional surveys showed that the proportion of self-reported White participants increased significantly and a sudden drop in participants self-identifying from other racial groups. (B-E) represents the mean difference of number of skipped questions between demographic groups specific to each survey for: (b) *Race*, (c)*: Ethnicity*, (d)*: Age*, (e)*: Sex at birth*. The bar plot with the error bar represents the bootstrap (1000 iterations) 95% confidence interval for pairwise group mean differences. (f): Race, (g): Ethnicity, (h): Age and (i): Sex at birth) represent pairwise statistically significant differences (*P*<.001) across all the self-reported demographic groups of the participants responded to the additional surveys. AA: African American; AoU: All of Us. HCA: Health Care Access. **P*<.001.

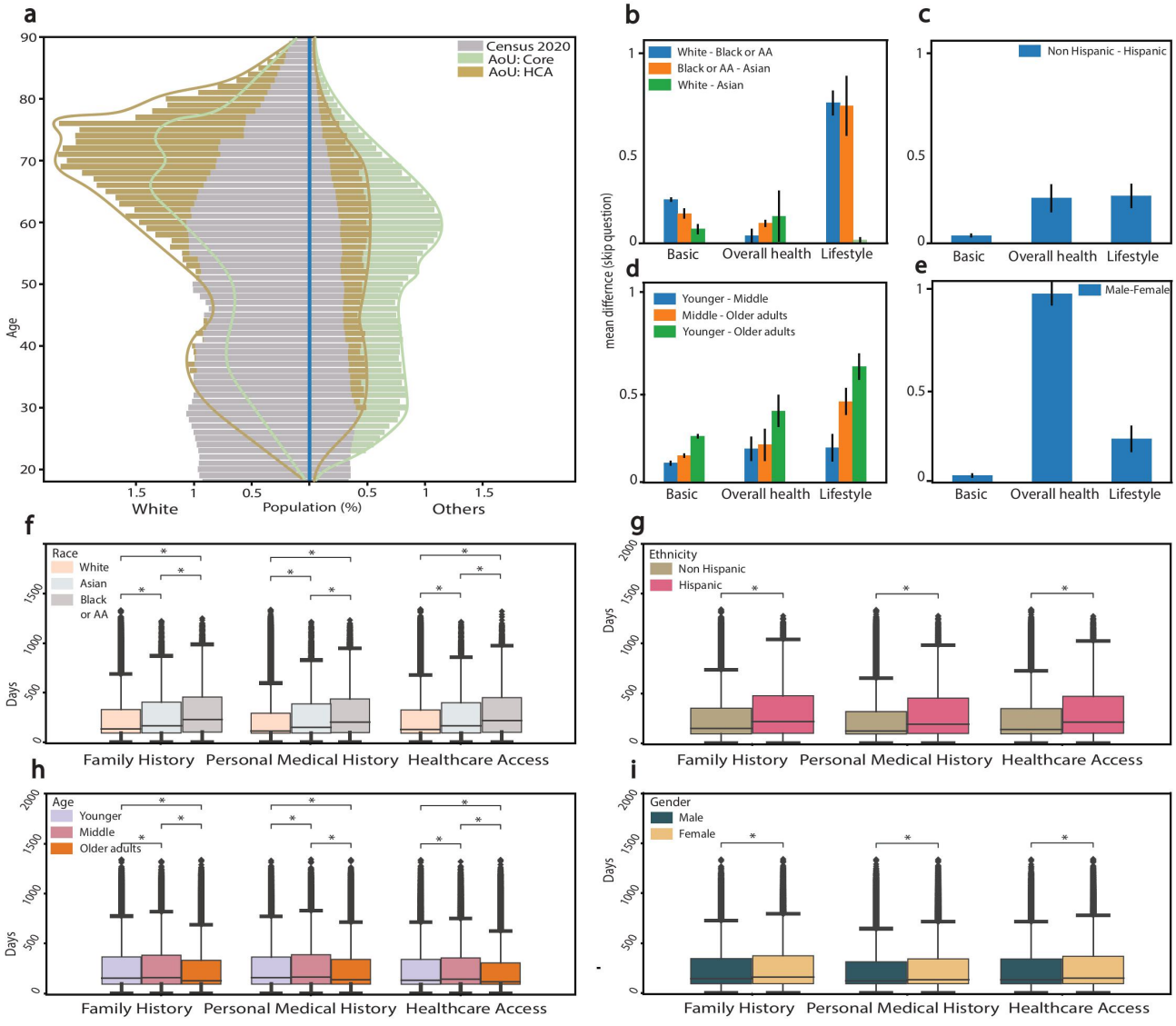


Table . Demographics characteristics and percentage distribution of participants in each component.

Groups	Core survey[a] (n=329,038), n (%)	HCA[b] survey (n=124,528), n (%)	Family history (n=117,693), n (%)	Medical history (n=113,830), n (%)	EHR[c] (n=271,421), n (%)	Covid survey (n=99,701), n (%)	Fitbit (n=11,615), n (%)
Age (years)							
18 - 44	108,254 (32.9)	35,740 (28.7)	33,307 (28.3)	32,328 (28.4)	88,212 (32.5)	23,130 (23.2)	3,984 (34.3)
45 - 64	121,415 (36.9)	42,090 (33.8)	39,545 (33.6)	38,019 (33.4)	100,697 (37.1)	33,698 (33.8)	4,321 (37.2)
≥65	99,369 (30.2)	46,698 (37.5)	44,841 (38.1)	43,483 (38.2)	82,512 (30.4)	42,971 (43.0)	3310 (28.6)
Sex							
Male	125,034 (38)	42,090 (33.8)	39,663 (33.7)	38,475 (33.8)	104,497 (38.5)	33,699 (33.8)	3,357 (28.9)
Female	199,726 (60.7)	81,441 (65.4)	77,207 (65.6)	74,559 (65.5)	163,395 (60.2)	65,304 (65.5)	8,200 (70.6)
Race							
White	177,681 (54)	93,645 (75.2)	89,094 (75.7)	86,625 (76.1)	144,124 (53.1)	78,963 (79.2)	9,698 (83.5)
African American	69,098 (21)	10,958 (8.8)	10,004 (8.5)	9,448 (8.3)	58,898 (21.7)	7,278 (7.3)	581 (5)
Asian	11,187 (3.4)	4,234 (3.4)	4,002 (3.4)	3,870 (3.4)	8,685 (3.2)	2,892 (2.9)	372 (3.2)
Ethnicity							
NonHispanic	258,953 (78.7)	109,336 (87.8)	103,570 (88)	100,398 (88.2)	212,523 (78.3)	89,631 (89.9)	10,674 (91.9)
Hispanic	60,543 (18.4)	12,577 (10.1)	11,652 (9.9)	11,041 (9.7)	51,027 (18.8)	7,876 (7.9)	767 (6.6)
Income (USD)							
<50,000	139,183 (42.3)	38,106 (30.6)	35,661 (30.3)	34,035 (29.9)	116,982 (43.1)	27,916 (28)	2,532 (21.8)
50,000-100,000	59,556 (18.1)	32,626 (26.2)	31,071 (26.4)	30,279 (26.6)	48,041 (17.7)	27,518 (27.6)	3,554 (30.6)
100,000-200,000	46,394 (14.1)	28,766 (23.1)	27,422 (23.3)	26,750 (23.5)	36,913 (13.6)	24,526 (24.6)	3,520 (30.3)
>200,000	20,071 (6.1)	12,702 (10.2)	12,005 (10.2)	11,724 (10.3)	15,851 (5.84)	10,469 (10.5)	1,371 (11.8)

[a]Core Survey at enrollment.

[b]HCA: Health Care Access.

[c]EHR: electronic health records.

Differences in Engagement

The relative shift in the proportions of participants is reported as percentage changes. We observed an increase of 21.2% in the proportion of self-identified White participants and correspondingly a decrease of Black or African American respondents by 12.18% in the HCA compared to the Core surveys. Similarly, the proportion of participants who identified as NonHispanic (of all races) increased by 9.1%, to 87.8% (n=109,336) of the total respondent pool in the HCA survey, while people self-identifying as Hispanic decreased by 8.33% to 10.1% of the respondent pool (Table 1 and Figure 2A). To explore the engagement further, we compared the percentage of each self-identified demographic group that continued in each optional component of the study (ie, optional surveys, EHR consent, COVID Surveys, and BYOD Fitbit) and observed that the engagement varied across most components except the BYOD Fitbit (Table 2). The engagement with optional surveys

was 3 times higher among people who identified as White, with 52.7% (n=93,614) of White participants who enrolled and also completed the optional surveys (HCA), compared to Black or African American participants (n=10,887, 15.8%), and nearly twice as high as those identifying as Asian (n=12531, 38.7%) (*P*<.001). Similarly, 42.2% (n=109,279) of those self-identifying as nonHispanic (of all races) engaged with optional surveys as compared to 20.7% (n=12,531) of self-identified Hispanic participants (*P*<.006). 63.4% (n=12,681) of participants who self-identified with high household income engaged in the optional surveys compared to 27.4% (n=38,190) of self-identified with low household income (*P*<.001). A total of 47% (n=46,743) of older adults (aged >65 years) engaged with optional surveys compared to 34.64% (n=37,504) and 33.05% (n=40,096)of the middle-aged and younger groups, respectively. Engagement was 40.8% (n=81,460) for participants who self-identified as female at birth engaged in the optional surveys compared to 33.66% (n=42,137) of self-identified male at birth. However, neither self-reported age nor sex at birth were statistically significant indicators of engagement using the χ^2 test.

The first COVID survey showed large differences in participation across age (*P*=.02), race (*P*<.001), household income (*P*<.001), and ethnicity (*P*<.001), but not by sex at birth (*P*=.45), Older individuals participated at higher rates, with 43.2% (n=42,964) of self-reported older adults compared to 27.8% (30,099) of middle-aged and 21.3% (n=25,841) of younger participants completing the first survey. Self-reported White participants (n=79,048, 44.5%) participated at more than 4 times higher levels than self-identified Black or African American participants (n=7281, 10.54%) and nearly twice those identifying as Asian (n=2880, 26.09%). Self-identified high-income participants (n=10,485, 52.4%) participated at more than approximately 2.5 times higher levels than self-identified low household income participants (n=27,876, 20%). Similarly, participants self-identifying as nonHispanic (n=89,552, 34.59%) engaged at almost 3 times higher rates than those identifying as Hispanic (n=7876, 13.01%).

We did not observe any statistically significant differences in engagement by self-reported age, race, ethnicity, or sex in participation rates for the BYOD Fitbit substudy or the consent to share EHR data.

Table . Demographics characteristics and response rate proportion specific to each survey in comparison to core surveys, (ie, response rate percentage = Number of participant responded to optional surveys/Number of participants responded to core survey)×100.

Groups	Core[a] (n)	HCA[b] survey, (%)	*P* value	Family history (%)	*P* value	Medical history, n (%)	*P* value	EHR[c], n (%)	*P* value	Covid survey, n (%)	*P* value	Fitbit[d], n (%)	*P* value
Age groups (years)			.22		.18		.19		.94		.02		.99
18 - 44	121,317	40,156 (33.1)		37,244 (30.7)		36,274 (29.9)		96,932 (79.9)		25,841 (21.3)		4,489 (3.7)	
45 - 64	108,268	37,461 (34.6)		35,295 (32.6)		33,996 (31.4)		91,162 (84.2)		30,099 (27.8)		3,898 (3.6)	
>65	99,453	46,743 (47)		44,853 (45.1)		43,461 (43.7)		82,745 (83.2)		42,964 (43.2)		3,282 (3.3)	
Sex			.41		.41		.42		.81		.45		.59
Male	125,184	42,187 (33.7)		39,683 (31.7)		38,431 (30.7)		105,530 (84.3)		33,674 (26.9)		3,380 (2.7)	
Female	199,658	81,460 (40.8)		77,068 (38.6)		74,672 (37.4)		162,122 (81.2)		65,288 (32.7)		8,186 (4.1)	
Race			<.001		<.001		<.001		0.40		<.001		.18
White	177,637	93,615 (52.7)		89,174 (50.2)		86,509 (48.7)		135,359 (76.2)		79,048 (44.5)		9,770 (5.5)	
African American	69,084	10,915 (15.8)		10,017 (14.5)		9,395 (13.6)		63,557 (92)		7,254 (10.5)		553 (0.8)	
Asian	11,040	4,272 (38.7)		3,974 (36)		3,875 (35.1)		8,578 (77.7)		2,881 (26.1)		364 (3.3)	
Ethnicity			.006		.006		.006		.41		.001		.23
NonHispanic	258,895	109,254 (42.2)		103,558 (40)		100,451 (38.8)		208,152 (80.4)		89,578 (34.6)		10,615 (4.1)	
Hispanic	60,535	12,531 (20.7)		11,623 (19.2)		11,078 (18.3)		55,147 (91.1)		7,870 (13)		787 (1.3)	
Income (USD)			.001		.001		.001		.98		<.001		.31
<50,000	139,380	38,162 (27.4)		35,639 (25.6)		34,092 (24.5)		134,125 (96.2)		27,946 (20.1)		2,537 (1.8)	
50,000-100,000	59,563	32,700 (54.9)		31,134 (52.3)		30,222 (50.7)		55,078 (92.5)		27,554 (46.3)		3,556 (6)	
100,000-200,000	46,381	28,728 (61.9)		27,411 (59.1)		26,729 (57.6)		42,346 (91.3)		24,512 (52.9)		3,520 (7.6)	
>200,000	20,008	12,685 (63.4)		12,043 (60.2)		11,785 (58.9)		18,181 (90.9)		10,494 (52.5)		1,377 (6.9)	

[a]Core survey at enrollment.

[b]HCA: Health Care Access.

[c]EHR: electronic health records.

[d]Fitbit was launched later in the study (see Figure 1). For a comparison of engagement rates between participants who joined before and after the launch of the Fitbit component (see Table S5 in Multimedia Appendix 1).

Response Time to Optional Surveys

In addition to differences in engagement, we also analyzed how long it took for participants to participate in optional surveys as measured by the time in days between enrollment and completion of the optional component. Given that only the three optional surveys (ie, HCA, PMH, and FMH) were available,

and the COVID Surveys and BYOD Fitbit, were unavailable at the launch of the study, we were only able to explore the delay for these surveys. We found differences in the response time based on self-reported age, sex at birth, race, ethnicity, and household income (all *Ps*<.001 based on Kruskal-Wallis one way analysis of variance). These results are summarized in Table S2 in Multimedia Appendix 1. Overall for the optional surveys, response times ranged from 0 to 1342 days with a mean of 247.9 (SD 233) days and median value was 141 (IQR 314.55) days. Self-reported older adults (median 117, IQR 213) took less time to respond to optional surveys compared to younger (median 135, IQR 251) and middle aged (median 138, IQR259) participants. People who identified as male at birth (median 120, IQR 221) took less time compared to people who self-identified as female at birth (median 133, IQR249). People who self-identified as White completed surveys sooner after enrollment (median 115, IQR 202) in comparison to those who self-identified as Black or African American (median 202, IQR 340), and Asian (median 153, IQR 295). People who identified as nonHispanic of any race (median 123, IQR 225) responded sooner to optional surveys as compared to people who self-identified as Hispanic (of all races) (median 188, IQR 335). People self-identified with high household income (median 119, IQR 195) took less time compared to people with low household income (median 150, IQR 279). We performed a secondary analysis to determine pairwise differences using the nonparametric Mann-Whitney test to illustrate which groups had quicker participation (Figure 2F–I). The linear mixed effect results presented in Table 3 provided valuable insights into the relationships between various demographic variables and their impact on response time. The table displays the estimated effects in terms of coefficients and relative percentage changes, with the use of bootstrap for accuracy. When examining age groups, individuals aged 45-64 years exhibited a 3.04% increase in response time compared to the reference group (18-44 years), while those aged ≥65 years displayed a decrease of −2.76%. Gender differences were also notable, with women showing a 5.65% increase in response time compared to men. In terms of race, African Americans exhibited a substantial 33.5% increase in response time compared to White individuals, while Asians and other racial groups also displayed significant differences. Ethnicity played a role as well, with Hispanics experiencing an 18.41% increase (Table 3) . Finally, income levels showed a consistent trend, with higher income brackets associated with decreased response times, highlighting the influence of socioeconomic factors on response behavior. These results offer important insights into the complex interplay between demographics and response times, shedding light on potential areas for further investigation and intervention. We conducted a secondary analysis to check if outliers were responsible for any significant difference and found no difference in the demographics. Distribution across demographic data is shown in Table S3 in Multimedia Appendix 1. We also explored the interaction effect (see Table S4 in Multimedia Appendix 1) between other demographic features and household income. The study found that the effect of age, gender, race, and ethnicity on response time was moderated by income. The response time was slower for people with higher income in all age groups, except 18-44 years. The difference was most pronounced for people aged ≥65 years. The interaction effect between gender and income was only significant for people who identify as neither male nor female. The interaction effect between race and income was significant for all race groups except White. The interaction effect between ethnicity and income was only significant for people who were not in plurality.

Table . Estimated effect (coefficient and relative % change using bootstrap) on response time.

Variables	Model	
	Coefficient	Percent change (95% CI)
Age (years)		
18-44	0	0
45-64	0.03	3.04 (2.99-3.09)
>65	–0.028	–2.76 (–2.78 to –2.69)
Gender		
Male	0	0
Female	0.055	5.65 (5.55-5.65)
Race		
White	0	0
African American	0.289	33.50 (33.40-33.57)
Asian	0.169	18.41 (18.30-18.54)
Ethnicity		
NonHispanic	0	0
Hispanic	0.169	18.41 (18.26-18.49)
Income (USD)		
<50,000	0	0
50,000-100,000	–0.045	–4.79 (–4.91 to –4.56)
100,000-200,000	–0.066	–6.39 (–6.45 to –6.31)
>200,000	–0.05	–4.88 (–4.96 to –4.78)

Completeness of Responses

To evaluate the frequency of skipped questions in each survey (survey completeness), we counted skipped questions and compared those across self-reported demographic groups (Table S5 in Multimedia Appendix 1). Skipped questions were relatively rare, with approximately 124,000 skipped questions out of 34 million answered; on average 329,038 participants skipped a mean of 1.32 (SD 0.74) questions per survey. We reported the results of "completeness of response" based on core surveys only, given that all the questions in the optional surveys were already answered except for one participant, who skipped one question. Using a nonparametric test, we observed differences in completeness for all self-reported demographic groups in the three core surveys. Mean values are reported for missing data (ie, skipped questions). Middle-aged participants skipped more questions (1.72 questions) compared to 1.54 for younger and 1.6 for seniors (P<.001). Participants who identified as female at birth skipped fewer questions (1.58) compared to participants who identified male at birth (1.99) (P<.001). Participants who self-reported as Black or African American race (1.86) and people self-reporting nonHispanic ethnicity of any race (1.71) skipped more questions compared to people of self-identified White (1.62) and Asian (1.62) race, and those who self-reported Hispanic (1.57) ethnicity of any race (P<.001). Participants with high household income skipped fewer questions (1.58) compared to participants with low household income (1.61) (P<.001). Also, we conducted pairwise group mean differences to further estimate the effect size which returned significant results (Figure 2B-E).

Secondary Analysis on Effect of Timing of Enrollment on Engagement

As all protocols were not launched at the same time, time of enrollment may have affected the engagement behavior of the participants who joined before and after the launch of the specific study components. We conducted a secondary analysis to evaluate if there were any differences in engagement between individuals who enrolled before and after the launch of BYOD Fitbit and Sync4Science separately. We found no significant differences in the demographic groups. Distribution across demographic variables for pre- and postlaunch of BYOD Fitbit (Table S6 in Multimedia Appendix 1), and for Sync4Science is shown in Table S7 in Multimedia Appendix 1.

Discussion

Principal Finding

The utility and translatability of studies is dependent on the study being carried out on a representative population. The AoURP has enrolled an extremely large cohort that aims to be representative of the diversity of the broader US population [25]. However, the data collected over time is not as representative as the enrolled population, with differences in engagement across demographic groups. This pattern is observed in both, the demographics that remained engaged and the degree

to which participants completed optional substudy components and how quickly they responded [26]. These variations are driven by socioeconomic, cultural, and structural factors, underscoring the need for tailored strategies to mitigate participation disparities [27]. Addressing these disparities is essential to maintain the cohort's representativeness, which directly impacts the validity and generalizability of the findings [28].

Comparison to Prior Work

This study builds on prior research by highlighting demographic engagement disparities in longitudinal cohort studies, an issue that has been well-documented in other population-based research efforts [29,30]. Previous studies have reported similar challenges, such as the underrepresentation of minority groups and the overrepresentation of higher socioeconomic status in voluntary participation metrics [31]. However, unlike earlier research, this study leverages the unique scale and diversity of the AoURP cohort to provide deeper insights into how engagement patterns vary over time and across demographic groups. By emphasizing socioeconomic and structural drivers of engagement disparities, this work highlights the need for target interventions necessary for retaining participants and ensuring equitable representation [32].

Strengths and Limitations

Our study benefits from a large, diverse cohort and longitudinal data, providing valuable insights into participation patterns. However, there are limitations that must be acknowledged. First, there may be biases due to self-selection and varying dropout rates among demographic groups. Overrepresentation of more engaged participants with higher socioeconomic status in optional surveys could skew health outcome estimates [7]. Second, the high dropout rates observed among underrepresented groups over time pose a challenge to maintaining sample diversity, which could limit the generalizability of findings to the broader US population [33]. Third, reliance on self-reported data, which may introduce reporting biases that affect the accuracy of the findings [27]. To address these limitations, efforts should focus on implementing strategies such as over-recruiting populations with higher attrition rates, developing tailored retention programs, and validating self-reported data through objective measures such as linked medical records [29]. These steps can enhance the reliability and applicability of the study's findings, ensuring their broader impact.

Future Directions

Based on our findings, decentralized studies such as AoURP should prioritize interventions that address engagement disparities. Over-recruiting populations with higher attrition rates and implementing targeted outreach strategies for underrepresented groups can enhance participation. Importantly, these strategies should be codesigned with input from participants to address specific barriers effectively [30]. Addressing challenges such as the digital divide, cultural differences, and logistical barriers is critical to fostering equity in research participation [31,34]. Innovative approaches such as bidirectional engagement where participants actively contribute to study design and receive meaningful feedback could further strengthen retention. Additionally, leveraging digital health tools and adaptive technologies can facilitate participant engagement while addressing socioeconomic barriers [35].

Conclusion

This study underscores the importance of maintaining representativeness in large, decentralized cohort studies such as AoURP. By identifying participation disparities and their underlying drivers, we provide actionable insights to improve cohort retention and engagement strategies. Future research should continue to explore and validate these interventions to ensure equitable and impactful scientific discoveries that benefit all segments of the population.

Acknowledgments

Research reported here was supported by the Office of the Director of the National Institutes of Health under award number U24OD023176. This work would not be possible without the partnership of the participants in the AoURP. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

Data are available only to researchers through the All of Us cloud-based Research Workbench. The complete code used for the analysis is available through a GitHub code repository [36].

Conflicts of Interest

All authors listed are employed at Sage Bionetworks. AP is an employee of Boehringer Ingelheim, USA.

Multimedia Appendix 1
Demographic characteristics.
[DOCX File, 199 KB - i-jmr_v14i1e56803_app1.docx]

References

1. Ford JG, Howerton MW, Lai GY, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer 2008 Jan 15;112(2):228-242. [doi: 10.1002/cncr.23157] [Medline: 18008363]
2. Bolen S, Tilburt J, Baffi C, et al. Defining "success" in recruitment of underrepresented populations to cancer clinical trials: moving toward a more consistent approach. Cancer 2006 Mar 15;106(6):1197-1204. [doi: 10.1002/cncr.21745] [Medline: 16453333]
3. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research. NIH Grants and Funding. 2023 Oct 30. URL: http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm [accessed 2025-02-28]
4. Food and drug administration modernization act (FDAMA) of 1997. US Food and Drug Administration. 2018 Nov 3. URL: https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/food-and-drug-administration-modernization-act-fdama-1997 [accessed 2023-10-30]
5. Medicare clinical trial policies. Centers for Medicare & Medicaid Services. 2023 Oct 30. URL: https://www.cms.gov/medicare/coverage/approved-facilities-trials-registries/clinical-policies?redirect=/ClinicalTrialPolicies [accessed 2025-02-28]
6. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]
7. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
8. Kolovson S, Pratap A, Duffy J, Allred R, Munson SA, Areán PA. Understanding participant needs for engagement and attitudes towards passive sensing in remote digital health studies. Int Conf Pervasive Comput Technol Healthc 2020 May;2020:347-362. [doi: 10.1145/3421937.3422025] [Medline: 33717638]
9. The All of Us Research Program Investigators. The "All of Us" Research Program. N Engl J Med 2019:668-676. [doi: 10.1056/NEJMsr1809937]
10. Who can join? All of Us Research Program. 2024 Jul 15. URL: https://www.joinallofus.org/who-can-join [accessed 2025-02-28]
11. All of us research program protocol summary. All of Us Research Program. 2024 Jul 15. URL: https://allofus.nih.gov/sites/default/files/All_of_Us_Protocol_Overview_Mar_2022.pdf [accessed 2025-02-28]
12. Survey explorer. All of Us Research Hub. 2022 Apr 25. URL: https://www.researchallofus.org/data-tools/survey-explorer/ [accessed 2025-02-28]
13. Researcher workbench. All of Us Research Hub. 2022 Feb 14. URL: https://www.researchallofus.org/data-tools/workbench/ [accessed 2025-02-28]
14. Helping patients share EHR data with researchers. Sync For Science. 2024 Apr 25. URL: http://syncfor.science/ [accessed 2025-02-28]
15. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995 Jan 21;310(6973):170. [doi: 10.1136/bmj.310.6973.170] [Medline: 7833759]
16. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
17. Statannotations 0.7.1. PYPI. URL: https://pypi.org/project/statannotations/ [accessed 2025-02-28]
18. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947 Mar;18(1):50-60. [doi: 10.1214/aoms/1177730491]
19. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin 1945 Dec;1(6):80. [doi: 10.2307/3001968]
20. Efron B, Tibshirani RJ. An introduction to the bootstrap: CRC Press; 1994.
21. All of Us institutional review board (IRB). All of Us Research Program | NIH. 2022 Apr 8. URL: https://allofus.nih.gov/about/who-we-are/institutional-review-board-irb-of-all-of-us-research-program [accessed 2025-11-25]
22. Do I need my project reviewed by the all of us institutional review board (IRB) in order to access this data using the researcher workbench? All of Us Research Hub. 2024 Nov 25. URL: https://www.researchallofus.org/faq/do-i-need-my-project-reviewed-by-the-all-of-us-institutional-review-board-irb-in-order-to-access-this-data-using-the-researcher-workbench/ [accessed 2025-02-25]
23. Release notes. All of Us Research Program. URL: https://support.researchallofus.org/hc/en-us/sections/360006773231-Release-Notes [accessed 2025-02-25]
24. Data: try out our new way to explore data. US census bureau. 2022 Apr 29. URL: https://www.census.gov/data [accessed 2025-02-28]
25. All of us research program. All of Us Research Program | NIH. 2020 Jun 1. URL: https://allofus.nih.gov/future-health-begins-all-us [accessed 2024-11-25]
26. Diversity and inclusion. All of Us Research Program | NIH. 2020 Jun 24. URL: https://allofus.nih.gov/about/diversity-and-inclusion [accessed 2025-11-25]
27. Ochieng CA, Minion JT, Turner A, Blell M, Murtagh MJ. What does engagement mean to participants in longitudinal cohort studies? A qualitative study. BMC Med Ethics 2021 Jun 24;22(1):77. [doi: 10.1186/s12910-021-00648-w] [Medline: 34167521]

28. Mapes BM, Foster CS, Kusnoor SV, et al. Diversity and inclusion for the All of Us research program: a scoping review. PLoS ONE 2020;15(7):e0234962. [doi: 10.1371/journal.pone.0234962] [Medline: 32609747]
29. Teague S, Youssef GJ, Macdonald JA, et al. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. BMC Med Res Methodol 2018 Nov 26;18(1):151. [doi: 10.1186/s12874-018-0586-7] [Medline: 30477443]
30. Costello L, Dare J, Dontje M, Lambert C, Straker L. Applying the 4Ps of social marketing to retain and engage participants in longitudinal cohort studies: generation 2 Raine study participant perspectives. BMC Med Res Methodol 2022 Nov 5;22(1):288. [doi: 10.1186/s12874-022-01778-4] [Medline: 36335312]
31. Murray AL, Xie T. Engaging adolescents in contemporary longitudinal health research: strategies for promoting participation and retention. J Adolesc Health 2024 Jan;74(1):9-17. [doi: 10.1016/j.jadohealth.2023.06.032] [Medline: 37690009]
32. Ferris EB, Wyka K, Evenson KR, et al. Recruitment and retention strategies for community-based longitudinal studies in diverse urban neighborhoods. JMIR Form Res 2021 Mar 24;5(3):e18591. [doi: 10.2196/18591] [Medline: 33759799]
33. Population studies: diverse cohorts (milestone 1.C). NIH Institute of Aging. URL: https://www.nia.nih.gov/research/milestones/epidemiology-population-studies/population-studies-diverse-cohorts [accessed 2025-02-28]
34. Durocher K, Boparai N, Jankowicz D, Strudwick G. Identifying technology industry-led initiatives to address digital health equity. Digit Health 2021;7:20552076211056156. [doi: 10.1177/20552076211056156] [Medline: 35340562]
35. Participant engagement and health equity workshop. National Institutes of Health (NIH). 2015 Sep 3. URL: https://www.nih.gov/allofus-research-program/participant-engagement-health-equity-workshop [accessed 2024-11-25]
36. Sage bionetworks. Github. URL: https://github.com/Sage-Bionetworks/allofus_engagement [accessed 2025-02-25]

Abbreviations

AoU: All of Us
AoURP: All of Us Research Program
BYOD: bring your own device
EHR: electronic health record
FMH: Family Medical History
HCA: Health Care Access
IRB: institutional review board
LMM: linear mixed model
PMH: Personal Medical History
UBR: underrepresented in biomedical research

Edited by TDA Cardoso; submitted 26.01.24; peer-reviewed by A Hassan, F Albalwy, V Nagesh; revised version received 02.01.25; accepted 17.01.25; published 19.03.25.

Please cite as:
Yadav V, Neto EC, Doerr M, Pratap A, Omberg L
Long-Term Engagement of Diverse Study Cohorts in Decentralized Research: Longitudinal Analysis of "All of Us" Research Program Data
Interact J Med Res 2025;14:e56803
URL: https://www.i-jmr.org/2025/1/e56803
doi:10.2196/56803

Original Paper

Policy Spotlight Effects on Critical Time-Sensitive Diseases: Nationwide Retrospective Cohort Study on Taiwan's Hospital Emergency Capability Categorization Policy

Chih-Yuan Lin[1,2], MSc, MD, PhD; Chih-Ching Liu[3], PhD; Yu-Tung Huang[4], PhD; Yue-Chune Lee[2], DrPH

[1]Department of Neurology, Taipei City Hospital Linsen Chinese Medicine Branch, Taipei City, Taiwan

[2]Institute of Health and Welfare, College of Medicine, Yangming Campus, National Yang Ming Chiao Tung University, Taipei, Taiwan, Taipei City, Taiwan

[3]Department of Health Care Management, National Taipei University of Nursing and Health Science, Taipei, Taiwan

[4]National Center for Geriatrics and Welfare Research, National Health Research Institutes, Yunlin, Taiwan

Corresponding Author:
Yue-Chune Lee, DrPH
Institute of Health and Welfare, College of Medicine, Yangming Campus
National Yang Ming Chiao Tung University, Taipei, Taiwan
No.155, Sec.2, Linong Street
Taipei City, 112
Taiwan
Phone: 886 225916681 ext 1526
Fax: 886 225916681
Email: DAS73@tpech.gov.tw

Abstract

Background: Taiwan's categorization of hospital emergency capability (CHEC) policy is designed to regionalize and dispatch critical patients. The policy was designed in 2009 to improve the quality of emergency care for critical time-sensitive diseases (CTSDs). The CHEC policy primarily uses time-based quality surveillance indicators.

Objective: We aimed to investigate the impact of Taiwan's CHEC policy on CTSDs.

Methods: Using Taiwan's 2005 Longitudinal Health Insurance Database, this nationwide retrospective cohort study examined the CHEC policy's impact from 2005 to 2011. Propensity score matching and difference-in-differences analysis within a generalized estimating equation framework were used to compare pre- and postimplementation periods. The study focused on acute ischemic stroke (AIS), ST-segment elevation myocardial infarction (STEMI), septic shock, and major trauma. AIS and STEMI cases, monitored with time-based indicators, were evaluated for adherence to diagnostic and treatment guidelines as process quality measures. Mortality and medical use served as outcome indicators. Major trauma, with evolving guidelines and no time-based monitoring, acted as a control to test for policy spotlight effects.

Results: In our cohort of 9923 patients, refined through 1:1 propensity score matching, 5566 (56.09%) were male and were mostly older adults. Our analysis revealed that the CHEC policy effectively improved system efficiency and patient outcomes, resulting in significant reductions in medical orders (–7.29 items, 95% CI –10.09 to –4.48; *P*<.001), short-term mortality rates (–0.09%, 95% CI –0.17% to –0.02%; *P*=.01) and long-term mortality rates (–0.09%, 95% CI –0.15% to –0.04%; *P*=.001), and total medical expenses (–5328.35 points per case, 95% CI –10,387.10 to –269.60; *P*=.04), despite a modest increase in diagnostic fees (376.37 points, 95% CI 92.42-660.33; *P*=.01). The CHEC policy led to notable increases in diagnostic fees, major treatments, and medical orders for AIS and STEMI cases. For AIS cases, significant increases were observed in major treatments (β=0.77; 95% CI 0.21-1.33; *P*=.007) and medical orders (β=15.20; 95% CI 5.28-25.11; *P*=.003) compared to major trauma. In STEMI cases, diagnostic fees significantly increased (β=1983.75; 95% CI 84.28-3883.21; *P*=.04), while upward transfer rates significantly decreased (β=–0.59; 95% CI –1.18 to –0.001; *P*=.049). There were also trends toward increased major treatments (β=0.30; 95% CI –0.03 to 0.62, *P*=.07), medical orders (β=11.92; 95% CI –0.90 to 24.73; *P*=.07), and medical expenses (β=24,275.54; 95% CI –640.71 to 4,991,991.78; *P*=.06), although these were not statistically significant. In contrast, no significant changes were identified in process or outcome quality indicators for septic shock. These findings suggest policy spotlight effects, reflecting a greater emphasis on diseases directly prioritized under the CHEC policy.

Conclusions: The CHEC policy demonstrated the dual benefits of reducing costs and improving patient outcomes. We observed unintended consequences of policy spotlight effects, which led to a disproportionate improvement in guideline adherence and process quality for CTSDs with time-based surveillance indicators.

(Interact J Med Res 2025;14:e54651) doi:10.2196/54651

KEYWORDS

categorization of hospital emergency capability; quality; time-sensitive diseases; emergency care; difference-in-differences

Introduction

Background

The Agency for Healthcare Research and Quality proposed the concept of time-sensitive diseases [1] using scientific data to maintain up-to-date guidelines and launched the Get with the Guidelines campaign to establish it as the basis for surveillance indicators of process and outcome quality [2]. Critical time-sensitive diseases (CTSDs) refer to life-threatening illnesses or injuries that require immediate emergency care, where rapid intervention is paramount to mitigate morbidity and mortality [3]. The various guidelines for managing time-sensitive events emphasize the crucial importance of time. In the context of acute ischemic stroke (AIS), the "time is brain" [4] goal focuses on the timely reperfusion treatments, including intravenous thrombolysis and mechanical thrombectomy; in ST-segment elevation myocardial infarction (STEMI), the "time is muscle" goal focuses on early reperfusion [5]; in septic shock events, the "early goal" focuses on early resuscitation [6]; and in major trauma cases, the "golden hour" goal focuses on the window of opportunity in which patients can undergo rescue operations [7].

The American Medical Association issued the categorization of hospital emergency capability (CHEC) guidelines [8] to classify hospitals according to their emergency care capabilities, thereby regionalizing and providing emergency medical services with references for transport emergency patients to the nearest appropriate hospitals [9], aiming to reduce preventable deaths. Most studies investigating the effects of this categorization, designation, and regionalization policy reported positive findings [10,11]. However, these studies mainly focused on a single disease entity [10,11] or region [11]. The CHEC policy often implements rigid time-based surveillance indicators. These indicators can affect disease-specific guideline adherence in clinical practice because they may reshape the behaviors of emergency department (ED) medical providers [12]. This phenomenon is related to the so-called policy spotlight effects, which influence medical care providers' assessment of how others perceive them [13]. More specifically, the policy spotlight effects refer to the perception of medical care providers regarding how policy makers interpret surveillance indicators and adjust their process-related behaviors accordingly [14]. Current emergency care policies often use time-based criteria as process quality indicators, which may exacerbate the policy spotlight effects [13]; however, the unintended effects or safety concerns generated by these effects remain unclear. Therefore, our study targeted 4 CTSDs: AIS, STEMI, septic shock, and major trauma [7]. Our research hypothesizes that emergency care providers might inadvertently give more attention to diseases under active surveillance while potentially neglecting those not thoroughly incorporated in this observation. This focus might be based on their perception of observer expectations [15].

Objective

We aimed to examine the effects of the CHEC policy on process quality and outcomes for CTSDs, addressing three research questions: (1) How does the CHEC policy impact the quality of diagnosis, treatment, and outcomes for these diseases? (2) How do policy spotlight effects influence the prioritization of diseases under active surveillance and impact emergency care providers' behaviors in this context? and (3) What are the potential consequences of policy spotlight effects?

Methods

Setting, Study Design, and Data Source

Taiwan's National Health Insurance is a single-payer, compulsory social insurance system that primarily operates on a fee-for-service basis. This study is based on the National Health Insurance 2005 Longitudinal Health Insurance Database (LHID2005), which contains 1 million random cases, including medical records and hospital information, collected since 1995. The LHID2005 was validated to represent medical use, diagnosis and treatment process, and outcome quality for CTSDs [16].

This nationwide retrospective cohort study uses propensity score matching (PSM) and difference-in-differences (DID) analysis to evaluate the impact of the CHEC policy on CTSD care quality and outcomes. The CHEC policy was initiated in August 2009, which integrated 190 hospitals into a network focusing on acute conditions such as stroke, myocardial infarction, major trauma, and perinatal care [17]. We divided our analysis into 2 periods: before CHEC (August 1, 2005, to July 31, 2009) and after CHEC (August 1, 2009, to July 31, 2011). This division aimed to assess the CHEC policy's effects distinctly from the ED quality improvement plan introduced in 2012. Well-established guidelines exist for AIS, STEMI, and septic shock. In contrast, the guidelines for major trauma are continuously evolving due to the variability in injury mechanisms, locations, and severity. Moreover, AIS and STEMI events are stringently monitored under the CHEC policy with specific time-based quality indicators, whereas septic shock and major trauma events are not (Table 1).

Table 1. Critical time-sensitive diseases and categorization hospital emergency capability (CHEC) policy indicators in Taiwan [18].

Quality indicator	Acute ischemic stroke	ST-segment elevation MI[a]	Septic shock	Major trauma
Guidelines development	• Well developed	• Well developed	• Well developed	• Developing
Major diagnosis indicator	• Brain imaging (eg, CT[b] and MRI[c])	• EKG[d]	• Blood culture	• Image study
Major treatment indicator	• Intravenous TPA[e]	• PCI[f]	• Antibiotics	• Rescue operation
Guideline's major goal	• Early thrombolysis • Time is brain	• Early reperfusion • Time is muscle	• Goal-directed therapy • Early goal	• Rescue operation • (Golden hour)
Guideline's time-based criteria	• 60 min	• 90 min	• 3 to 6 h	• 1 h
CHEC policy indicators	• Stroke team • NIHSS[g] score evaluation • Intravenous TPA	• PCI team • Give aspirin and clopidogrel	• ICUS[h] critical care team	• Trauma team • ISSi evaluation
CHEC policy time-based criteria	• Neurologist consultation in <30 min • Door to CT in <30 min • Door to CT read in <45 min • Onset to needle in <3 h	• Cardiologist consultation in <30 min • Door to EKG in <10 min • Door to needle in <30 min • Door to balloon in <90 min	• Admission in <24 h	• Trauma team activation in <30 min

[a]MI: myocardial infarction.

[b]CT: computed tomography.

[c]MRI: magnetic resonance imaging.

[d]EKG: electrocardiography.

[e]TPA: tissue plasminogen activator.

[f]PCI: percutaneous coronary intervention.

[g]NIHSS: National Institute of Health Stroke Scale or Score.

[h]ICU: intensive care unit.

[i]ISS: Injury Severity Score.

Thus, we selected major trauma events as a reference for our study because they were not monitored under the CHEC policy with rigid indicators. We adopted pre- and postimplementation of a CHEC policy, using 1:1 PSM to control for confounding variables. We used the DID estimation approach to estimate the association of the CHEC policy on process and outcomes for AIS and STEMI. For the counterfactual, we used major trauma cases unexposed to the clinical guideline or CHEC policy time-based quality indicators as the basis for comparison.

Identification of Study Cohort

This study identified CTSDs based on ED visits accompanied by a primary diagnosis using the appropriate disease code. The identification of AIS (ie, codes 433 and 434), STEMI (ie, code 410), and septic shock (ie, codes 038, 785, and 995) was based on the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD - 9 - CM*). Major trauma cases were classified following the American Academy of Surgery Committee guidelines (ie, codes 800-959) [19]. Due to the absence of trauma severity data in the LHID2005 database, primary *ICD-9-CM* codes served as our initial method for identifying major trauma incidents. This identification was further refined by including cases where patients received rescue surgery or were admitted to the intensive care unit, serving as additional criteria for major trauma [20]. We excluded cases before the study period and those without hospital or patient sociodemographic information. We also excluded hospitals with a volume of <5 CTSD cases per year [21]. We used the date of the first ED visit as the index date.

Definition of Variables

The independent variable in this study was exposure to the CHEC policy intervention. Events related to AIS and STEMI were subject to rigid time-based quality indicators and regular surveillance under the CHEC policy. In contrast, despite having well-developed clinical guidelines, septic shock was not included under the CHEC policy's stringent time-based quality indicators. Similarly, major trauma cases, which lack well-developed clinical guidelines, were not subject to these policy indicators and were used as counterfactuals in this study. The dependent variables were divided into primary and secondary outcomes. Primary outcomes included guideline adherence to diagnostic and treatment process quality indicators. For diagnostic adherence, AIS was assessed by the completion of brain imaging (eg, computed tomography or magnetic resonance imaging) within 60 minutes of hospital arrival, while

STEMI was evaluated based on the completion of electrocardiography within 10 minutes. For treatment adherence, AIS required the administration of intravenous thrombolysis within 3 hours of symptom onset, and STEMI was assessed according to whether percutaneous coronary intervention was performed within 90 minutes of hospital arrival. Secondary outcomes included upward transfer rates, diagnostic fees, medical orders and expenses, and mortality rates. Upward transfer rates were defined as the proportion of patients transferred from lower-level hospitals to higher-level facilities. Diagnostic fees referred to the total costs incurred from diagnostic procedures during ED care. Medical orders and expenses represented the number and costs of medical interventions performed during ED visits. Mortality rates were measured as 30-day mortality, indicating the percentage of patients who died within 30 days from the index date (the emergency visit date), and 1-year mortality, representing deaths occurring within 1 year.

Covariates included patient-related predisposing factors, such as age, sex, and occupation. The enabling factor was the insured salary, while the Charlson Comorbidity Index, calculated using *ICD-9-CM* codes from primary diagnoses recorded in inpatient and outpatient claims data up to a year before the index date, served as a measure of health needs. Furthermore, external environmental factors, including urbanization and regional emergency resources, were considered. For hospital-level variables, the input-throughput-output model of Asplin et al [22] was used. Input was gauged using the rate of ED visits with Emergency Severity Index 1, while throughput and output efficiency were assessed via the ED's occupancy rate. This comprehensive framework ensured a robust evaluation of the CHEC policy's impact while accounting for potential confounding variables and contextual factors.

Statistical Analysis

Patients' characteristics, process quality, and outcomes were presented using descriptive statistics. Continuous data were described using mean (SD), and categorical data were presented using numbers and percentages. To enhance the robustness of the outcomes, we calculated the propensity score using a multivariable logistic regression that included all baseline covariates. The standardized mean difference was calculated to confirm the balance of potential confounders at baseline between groups before and after matching. A standardized mean difference of <0.1 was considered to represent a negligible difference [23].

We evaluated the impact of the CHEC intervention on each outcome, including overall differences, the differences within individual diseases, and between-disease differences in changes from baseline (ie, group-by-disease interaction effects), using a DID framework integrated with generalized estimating equation (GEE) models. The DID approach allowed us to compare changes in outcomes between the before and after CHEC policy periods across diseases, enabling the estimation of differential effects of the intervention while controlling for time-invariant unobserved confounders. This method is particularly suitable for evaluating policy interventions by focusing on within-group changes relative to a reference group over time. The GEE model is specified as follows:

$$Y_{ij} = \beta_0 + \beta_1 (\text{CHEC policy})_j + \beta_2 \text{ disease } + \beta_3 (\text{CHEC policy} \times \text{disease})_{ij} + \varepsilon_{ij,}$$

where Y_{ij} represents the outcome variable (eg, diagnosis indicator, treatment indicator, mortality, or medical use) for individual i at time j. β_0 is the intercept, capturing the baseline level of the outcome for the reference disease (ie, major trauma) in the pre-CHEC period. CHEC policy$_j$ is a binary variable (0=before CHEC, 1=after CHEC), and its coefficient β_1 captures the overall impact of the CHEC policy across all diseases. Disease$_i$ is a categorical variable representing the 4 diseases (ie, AIS, STEMI, septic shock, and major trauma), with β_2 estimating disease-specific differences at baseline. The interaction term (CHEC policy × disease)$_{ij}$, with coefficient β_3, reflects the differential impact of the CHEC policy for each disease compared to the reference group (ie, major trauma). A positive β_3 value indicates a greater change in the outcome for the respective disease relative to major trauma. All statistical analyses were performed using SAS (version 9.4; SAS Institute), and statistical significance was defined as value of <.05.

Ethical Considerations

This study used secondary data from the NIH LHID2005, which are fully anonymized and deidentified to protect participant privacy. The dataset contained no personal identifiers, such as names, addresses, or social security numbers, so individual informed consent was not required. The original consent provided for primary data collection, and the institutional review board approval covered secondary analyses without requiring additional consent. All analyses were conducted in compliance with relevant regulations to safeguard confidentiality, and data access was restricted to authorized researchers under institutional guidelines. The study received ethics approval from the Taiwan National Yang-Ming University Institutional Review Board (YM107035E).

Results

Participants' Characteristics

During the study period, we analyzed emergency presentations related to 4 CTSDs, originally encompassing 288,443 patients. Exclusion criteria included the diagnosis of CTSDs before 2005 (n=99,768, 34.59%), patients with transient ischemic attack or intracranial hemorrhage (n=878, 0.3%), non-STEMI cases (n=1315, 0.45%), individuals with major traumas defined by *ICD-9-CM* codes that did not necessitate a rescue operation or intensive care unit admission (n=142,446, 49.38%), and cases lacking hospital or living area information or where the hospital's volume of CTSDs was <5 visits per year (n=673, 0.23%). These criteria refined the total sample size to 43,363 (15.03%) patients. Considering the extended period before the policy intervention, this research adopted a 1:1 PSM technique, resulting in a final matched sample of 9923 patients. The flowchart and baseline table (Figure 1) display the initial count of emergency patients with CTSDs and the numbers after PSM, broken down by each of the 4 diseases. Table 2 presents the

PSM of participants with CTSDs before and after the PSM. After the matching process, each variable baseline characteristic demonstrated almost complete congruity. In addition, uniformity was achieved within each disease subgroup after matching (Table S1 in Multimedia Appendix 1). In 9923 patients, the distribution for each disease before and after PSM was as follows: AIS (n=2895, 29.17%), STEMI (n=723, 7.29%), septic shock (n=5441, 54.83%), and major trauma (n=864, 8.71%).

Septic shock was the most prevalent condition, accounting for 54.83% (5441/9923) of all cases. The patient population was male-dominated (5566/9923, 56.09%), with the majority (6084/9923, 61.31%) aged ≥65 years. Nearly three-fourths of the CTSD cases (7563/9923, 76.12%) were managed in hospitals categorized as moderate or severe levels. Care provided by specialty consultants accounted for 67.9% (6738/9923) of the cases.

Figure 1. Flow diagram of the study.

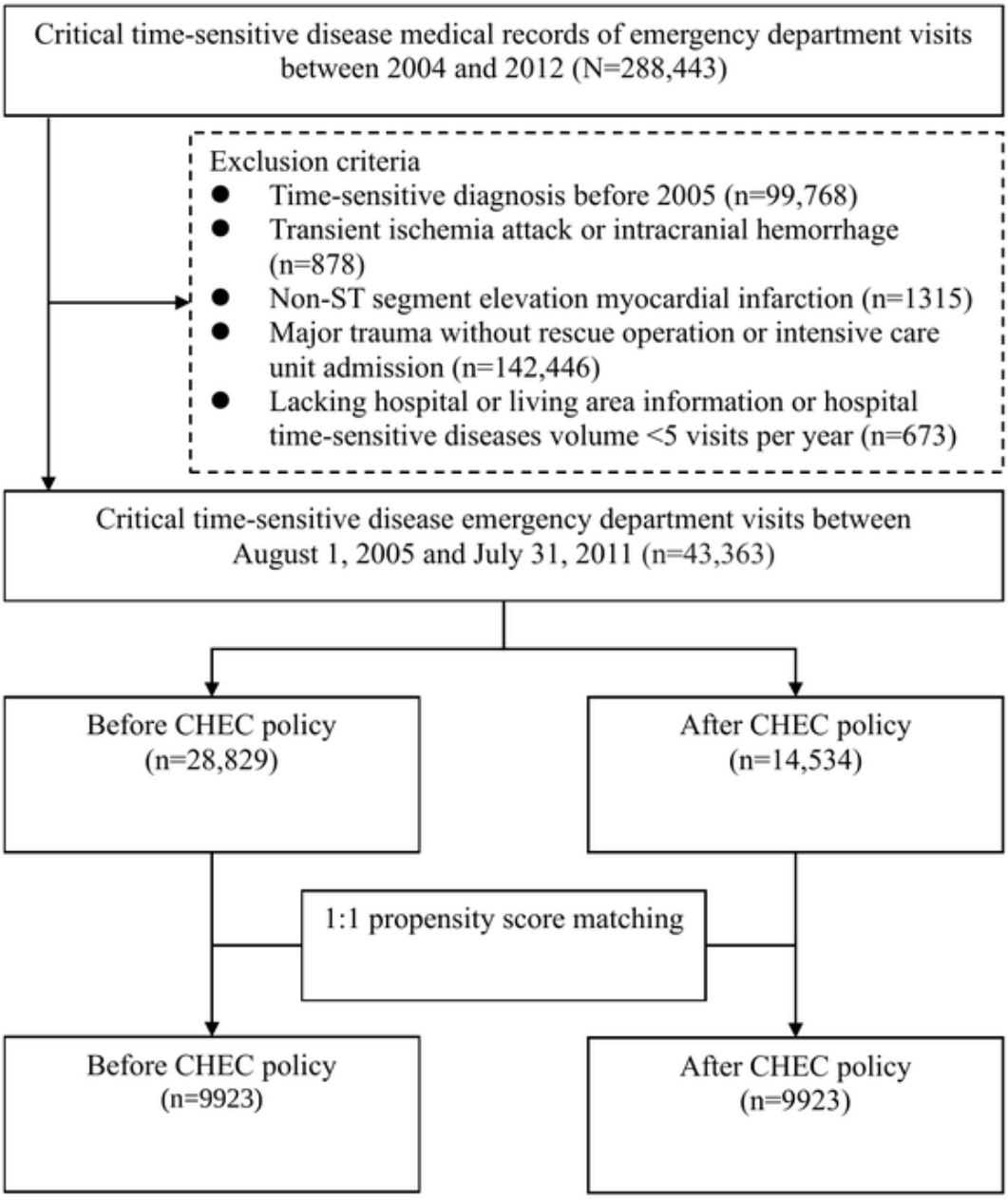


Table 2. Propensity score matched the comparison of patient characteristics before and after the categorization of hospital emergency capability (CHEC) policy. The complete table is available in Supplementary Table 2: Association of CHEC policy with process and outcome quality in four critical time-sensitive diseases (Multimedia Appendix 2).

Patient characteristics	Before propensity score matching			After propensity score matching		
	Before CHEC policy (n=28,829)	After CHEC policy (n=14,534)	ASMD[a]	Before CHEC policy (n=9923)	After CHEC policy (n=9923)	ASMD
Sex, n (%)						
Female	12,345 (42.82)	6250 (43)	0.004	4357 (43.91)	4268 (43.01)	0.018
Male	16,484 (57.18)	8284 (57)	0.004	5566 (56.09)	5655 (56.99)	0.018
Age (y), n (%)						
≤45	4105 (14.24)	1967 (13.53)	0.021	1289 (12.99)	1306 (13.16)	0.005
45-64	7173 (24.88)	3795 (26.11)	0.028	2550 (25.7)	2590 (26.1)	0.009
≥65	17,551 (60.88)	8772 (60.36)	0.011	6084 (61.31)	6027 (60.74)	0.012
Charlson Comorbidity Index, n (%)						
≤1	13,215 (45.84)	6546 (45.04)	0.016	4473 (45.08)	4434 (44.68)	0.008
>1	15,614 (54.16)	7988 (54.96)	0.016	5450 (54.92)	5489 (55.32)	0.008
Income[b] (NT $), n (%)						
≤22,800	15,961 (55.36)	6203 (42.68)	0.256	4428 (44.62)	4604 (46.4)	0.036
>22,800	12,868 (44.64)	8331 (57.32)	0.256	5495 (55.38)	5319 (53.6)	0.036
Occupation, n (%)						
Dependents of the insured	10,095 (35.02)	4952 (34.07)	0.020	3550 (35.78)	3326 (33.52)	0.048
Civil servants, teachers, and military personnel	2792 (9.68)	1529 (10.52)	0.028	986 (9.94)	1053 (10.61)	0.022
Nonmanual workers and professionals	2231 (7.74)	1240 (8.53)	0.029	748 (7.54)	796 (8.02)	0.018
Manual workers	10,250 (35.55)	5154 (35.46)	0.002	3422 (34.49)	3708 (37.37)	0.060
Others	3461 (12.01)	1659 (11.41)	0.019	1217 (12.26)	1040 (10.48)	0.056
Hospital categorization, n (%)						
Severe level	11,371 (39.44)	5924 (40.76)	0.027	3331 (33.57)	3502 (35.29)	0.036
Moderate level	10,921 (37.88)	5496 (37.81)	0.001	4030 (40.61)	4052 (40.83)	0.004
General level	6537 (22.68)	3114 (21.43)	0.030	2562 (25.82)	2369 (23.87)	0.045
ESI[c] triage level 1 and 2 rate, mean (SD)	5.8 (8.98)	16.63 (10.06)	1.136	14.01 (10.39)	14.37 (10.50)	0.034
Length of ED[d] stay, mean (SD)	5.33 (10.11)	4.49 (9.30)	0.086	6.25 (9.30)	5.36 (10.26)	0.091
ED observation ≥1-d rate, mean (SD)	11.37 (9.97)	12.32 (10.06)	0.095	10.53 (9.34)	11.32 (10.12)	0.081
Place of ED resources, n (%)						
Sufficiency	22,770 (78.98)	11,460 (78.85)	0.003	7768 (78.28)	7834 (78.95)	0.016
Not sufficiency	6059 (21.02)	3074 (21.15)	0.003	2155 (21.72)	2089 (21.05)	0.016
Time-sensitive disease, n (%)						
Acute ischemic stroke	8660 (30.04)	3814 (26.24)	0.085	2895 (29.17)	2895 (29.17)	0.000
ST-segment elevation MI[e]	2481 (8.61)	1141 (7.85)	0.028	723 (7.29)	723 (7.29)	0.000
Septic shock	14,896 (51.67)	8275 (56.94)	0.106	5441 (54.83)	5441 (54.83)	0.000
Major trauma	2792 (9.68)	1304 (8.97)	0.024	864 (8.71)	864 (8.71)	0.000

Patient characteristics	Before propensity score matching			After propensity score matching		
	Before CHEC policy (n=28,829)	After CHEC policy (n=14,534)	ASMD[a]	Before CHEC policy (n=9923)	After CHEC policy (n=9923)	ASMD
Delivery of care, n (%)						
Specialty consultant[f]	18,692 (64.84)	9245 (63.61)	0.026	6738 (67.9)	6570 (66.21)	0.036
Emergency physician	7788 (27.01)	4236 (29.15)	0.048	2449 (24.68)	2585 (26.05)	0.031
Others	2349 (8.15)	1053 (7.25)	0.034	736 (7.42)	768 (7.74)	0.012

[a]ASMD: absolute standardized mean difference.

[b]INT $1=US $0.03057 as of February 16, 2025.

[c]ESI: Emergency Severity Index.

[d]ED: emergency department.

[e]MI: myocardial infarction.

[f]Specialty consultant: (1) acute ischemic stroke is treated by neurologists, (2) acute MI is treated by cardiologists, (3) septic shock is managed by internal medicine physicians or critical care intensivists, and (4) major trauma conditions are managed by surgeons or critical care intensivists.

Impact of CHEC Policy Overall and on the Processes and Outcomes of the 4 Individual CTSDs Before and After Implementation

In examining individual diseases, primary diagnostic indicators for AIS, septic shock, and major trauma decreased after intervention, while only those for STEMI increased (Table 3).

Diagnostic fees increased for AIS, STEMI, and major trauma cases but decreased for septic shock cases. A similar trend was observed in the primary treatment indicators, which increased for AIS and STEMI cases and decreased for septic shock and major trauma cases. In contrast, medical orders showed a universal decline. Upward transfer rates increased for AIS and major trauma cases but decreased for STEMI and septic shock cases. Regarding outcome indicators, short-term and long-term mortality rates displayed a universal decline, except for AIS cases, which showed an increase. The medical expenses were higher for AIS and STEMI cases but lower for septic shock and major trauma cases.

In assessing the overall policy effects on 4 CTSD cohorts, the primary diagnosis indicator significantly decreased by 0.21% points (95% CI –0.29% to –0.13%; $P<.001$) and medical orders per case dropped by an average of 7.29 items (95% CI –10.09 to –4.48; $P<.001$). In comparison, diagnostic fees demonstrated an average increase of 376.37 points (95% CI 92.42-660.33; $P=.01$). The 30-day mortality rate saw a notable reduction of 0.09% points (95% CI –0.17% to –0.02%; $P=.01$), 1-year mortality significantly decreased by 0.09% points (95% CI –0.15% to –0.04%; $P=.001$), and medical expense per case significantly decreased by 5328.35 points (95% CI –10,387.10 to –269.60; $P=.04$).

Table 3. Comparative analysis of the individual and overall impact of categorization of hospital emergency capability (CHEC) policy effects on 4 critical time-sensitive diseases (N=9923).

Outcome	Change between before and after CHEC policy				Before CHEC policy	After CHEC policy	Multivariable model, β_1[a] (95% CI)	*P* value
	Acute ischemic stroke[b]	STEMI[c]	Septic shock[d]	Major trauma[e]				
Process quality								
Major diagnosis indicator, n (%)	–3.52	0.55	–2.28	–3.01	8682 (87.49)	8434 (84.99)	–0.21 (–0.29 to –0.13)	<.001
Diagnostic fees[f]	460.66	2746.6	–44.8	762.85	7166.94 (10,018.69)	7543.31 (10,833.26)	376.37 (92.42 to 660.33)	.01
Major treatment indicator, n (%)	0.62	2.07	–1.25	–3.12	4334 (43.68)	4272 (43.05)	–0.03 (–0.07 to 0.02)	.24
Medical orders per case	–0.93	–4.21	–9.67	–16.13	102.13 (109.19)	94.84 (101.07)	–7.29 (–10.09 to –4.48)	<.001
Upward transfer rate, n (%)	0.59	–1.24	–0.1	2.43	148 (1.49)	172 (1.73)	0.15 (–0.06 to 0.37)	.16
Outcomes								
30-d mortality, n (%)	–0.04	–1.52	–1.84	–1.27	1631 (16.44)	1508 (15.2)	–0.09 (–0.17 to –0.02)	.01
1-y mortality, n (%)	0.28	–0.83	–3.64	–0.58	3242 (32.67)	3041 (30.65)	–0.09 (–0.15 to –0.04)	.001
Total medical expense per case	3616.16	11,219.01	−11,059.09	−13,056.54	100,875.96 (192,912.94)	95,547.60 (176,487.79)	–5328.35 (–10,387.10 to –269.60)	.04

[a]Overall impact of the CHEC policy across all diseases.

[b]Acute ischemic stroke major diagnosis indicator: head image and major treatment indicator: intravenous tissue plasminogen activator thrombolysis.

[c]STEMI: ST-elevation myocardial infarction major diagnosis indicator: electrocardiography and major treatment indicator: percutaneous coronary intervention.

[d]Septic shock major diagnosis indicator: culture and major treatment indicator: antipathogen medication.

[e]Major trauma major diagnosis indicator: computed tomography, magnetic resonance imaging, or sonography study and major treatment indicator: rescue operation.

[f]Diagnostic fees: since Taiwan's National Health Insurance system operates on a global budget with reimbursement based on a point system, the actual monetary value of each point fluctuates. Currently, 1 National Health Insurance point is reimbursed at <NT $0.90 (US $0.0275) per point based on the latest exchange rate (1 NT $=US $0.03057 as of February 16, 2025).

Association of CHEC Policy With Processes and Outcome Quality in the 4 CTSDs

As presented in Table 4, model 1 examined the changes in indicators for individual diseases before and after the implementation of the CHEC policy, and the results show significant improvements in process quality measures. For AIS cases, following the implementation of the CHEC policy, there was a significant decrease in major diagnosis indicators by 0.23% points (95% CI –0.36% to –0.10%; *P*<.001). Conversely, the major treatment indicator experienced a significant increase of 0.57% points (95% CI 0.07%-1.07%; *P*=.03), and the upward transfer rate also significantly increased by 0.52% points (95% CI 0.02%-1.03%; *P*=.04). Moreover, there was a trend of increasing diagnostic fees, with a rise of 460.66 points (95% CI –3.44 to 924.76; *P*=.05). For STEMI cases, the diagnostic fees significantly increased by 2746.59 points (95% CI 1141.67-4351.51; *P*<.001). When examining septic shock, the major diagnosis indicator saw a significant decrease of 0.25% points (95% CI –0.37% to –0.12%; *P*<.001) following the introduction of the CHEC policy. Thirty-day mortality decreased by 0.11% (95% CI –0.20% to –0.02%; *P*=.02), and one-year mortality decreased by 0.15% (95% CI –0.22% to –0.07%; *P*<.001). In addition, medical orders significantly dropped by 9.67 items (95% CI –13.99 to –5.35; *P*<.001), and average medical expenses significantly decreased by 11,059.10 points (95% CI –18,603.60 to –3514.55; *P*=.004). Finally, regarding major trauma cases, after CHEC policy implementation, the average medical orders significantly decreased by 16.13 items (95% CI –25.32 to –6.94; *P*<.001).

Table 4. Association of categorization of hospital emergency capability (CHEC) policy with process and outcome quality in 4 critical time-sensitive diseases.

Critical time-sensitive disease	Change between before and after CHEC policy	Model 1[a]		Model 2[b]	
		β_2 (95% CI)	*P* value	β_3 (95% CI)	*P* value
Acute ischemic stroke[c] (n=2895)					
Major diagnosis indicator, n (%)	–3.52	–0.23 (–0.36 to –0.10)	<.001	–0.06 (–0.32 to 0.20)	.66
Diagnostic fees[d]	460.66	460.66 (–3.44 to 924.76)	.05	–302.19 (–1419.15 to 814.77)	.60
Major treatment indicator, n (%)	0.62	0.57 (0.07 to 1.07)	*.03*	0.77 (0.21 to 1.33)	.007
Medical orders per case	−0.93	–0.93 (–4.64 to 2.78)	.62	15.20 (5.28 to 25.11)	.003
Upward transfer rate, n (%)	0.59	0.52 (0.02 to 1.03)	.04	0.21 (–0.39 to 0.81)	.49
Short-term mortality (30 d), n (%)	–0.04	−0.01 (−0.25 to 0.23)	.95	0.11 (–0.27 to 0.49)	.57
Long-term mortality (365 d), n (%)	0.28	0.02 (−0.12 to 0.16)	.77	0.06 (–0.22 to.033)	.69
Total medical expense per case[h]	3616.16	3616.15 (–3524.26 to 10,756.56)	.32	16,672.69 (–3581.75 to 36,927.12)	.11
ST-segment elevation myocardial infarction[e](n=723)					
Major diagnosis indicator, n (%)	0.55	0.09 (–0.32 to 0.49)	.68	0.26 (–0.21 to 0.72)	.28
Diagnostic fees	2746.6	2746.59 (1141.67 to 4351.51)	<.001	1983.75 (84.28 to 3883.21)	*.04*
Major treatment indicator, n (%)	2.07	0.09 (–0.11 to 0.30)	.38	0.30 (–0.03 to 0.62)	.07
Medical orders per case	−4.21	–4.21 (–13.14 to 4.72)	.36	11.92 (–0.90 to 24.73)	.07
Upward transfer rate, n (%)	–1.24	–0.28 (–0.76 to 0.21)	.27	–0.59 (–1.18 to –0.001)	.049
Short-term mortality (30 d), n (%)	–1.52	–0.10 (–0.36 to 0.16)	.45	–0.02 (–0.37 to 0.41)	.92
Long-term mortality (365 d), n (%)	–0.83	−0.04 (−0.26 to 0.18)	.72	–0.01 (–0.33 to .032)	.97
Total medical expense per case	11,219.01	11,219.00 (–4953.90 to 27,391.90)	.17	24,275.54 (–640.71 to 49,191.78)	.06
Septic shock[f] (n=5441)					
Major diagnosis indicator, n (%)	–2.28	−0.25 (−0.37 to −0.12)	<.001	–0.08 (–0.33 to 0.18)	.56
Diagnostic fees	–44.8	–44.80 (–412.26 to 322.66)	.81	–807.65 (–1888.03 to 272.74)	.14
Major treatment indicator, n (%)	–1.25	−0.06 (−0.14 to 0.02)	.14	0.14 (–0.12 to 0.41)	.28
Medical orders per case	–9.67	–9.67 (–13.99 to –5.35)	<.001	6.45 (–3.70 to 16.61)	.21
Upward transfer rate, n (%)	–0.1	–0.27 (–0.93 to 0.38)	.41	–0.59 (–1.32 to 0.15)	.12
Short-term mortality (30 d), n (%)	–1.84	–0.11 (–0.20 to –0.02)	.02	–0.01 (–0.29 to 0.31)	.95
Long-term mortality (365 d), n (%)	–3.64	–0.15 (–0.22 to –0.07)	<.001	–0.11 (–0.36 to 0.13)	.36
Total medical expense per case	–11,059.09	–11,059.10 (–18,603.60 to –3514.55)	.004	1997.45 (–18,403.00 to 22,397.86)	.85
Major trauma[g] (n=864)					
Major diagnosis indicator, n (%)	–3.01	–0.17 (–0.39 to 0.05)	.14	N/A[h]	N/A
Diagnostic fees	762.85	762.85 (–253.13 to 1778.82)	.14	N/A	N/A
Major treatment indicator, n (%)	–3.12	–0.21 (–0.46 to 0.05)	.11	N/A	N/A
Medical orders per case	–16.13	–16.13 (–25.32 to –6.94)	<.001	N/A	N/A
Upward transfer rate, n (%)	2.43	0.31 (–0.02 to 0.65)	.06	N/A	N/A
Short-term mortality (30 d), n (%)	–1.27	–0.12 (–0.41 to 0.17)	.43	N/A	N/A
Long-term mortality (365 d), n (%)	–0.58	–0.03 (–0.27 to 0.20)	.77	N/A	N/A

Critical time-sensitive disease	Change between before and after CHEC policy	Model 1[a]		Model 2[b]	
		β_2 (95% CI)	*P* value	β_3 (95% CI)	*P* value
Total medical expense per case	−13,056.54	–13,056.50 (–32,010.60 to 5897.53)	.18	N/A	N/A

[a]Model 1 compares the specific disease differences between before and after CHEC policy implementation.

[b]Model 2: model-adjusted estimates for an interaction between a binary measure of CHEC policy (ie, postimplementation vs preimplementation) and critical time-sensitive diseases compared with major trauma (eg, acute ischemic stroke vs major trauma, ST-segment elevation myocardial infarction vs major trauma, and septic shock vs major trauma).

[c]Acute ischemic stroke major diagnosis indicator: head image and major treatment indicator: intravenous tissue plasminogen activator thrombolysis.

[d]Diagnostic fees: since Taiwan's National Health Insurance system operates on a global budget with reimbursement based on a point system, the actual monetary value of each point fluctuates. Currently, 1 National Health Insurance point is reimbursed at <NT $0.90 (US $0.0275) per point based on the latest exchange rate (NT $1=US $0.03057 as of February 16, 2025).

[e]STEMI: ST-segment elevation myocardial infarction major diagnosis indicator: electrocardiography and major treatment indicator: percutaneous coronary intervention.

[f]Septic shock major diagnosis indicator: culture and major treatment indicator: antipathogen medication.

[g]Major trauma major diagnosis indicator: computed tomography, magnetic resonance imaging, or sonography study and major treatment indicator: rescue operation.

[h]N/A: data not applicable as major trauma cases were the reference group.

In model 2, results from the GEE model highlighted the CHEC policy's varied effects across different diseases. A positive group-by-disease interaction β_3 coefficient indicated that the outcome changes for that disease was greater than the reference group (Table 4). Compared to major trauma cases, AIS cases exhibited a significant increase in the major treatment indicator (interaction β coefficient=0.77; 95% CI 0.21-1.33; *P*=.007) and medical orders (interaction β coefficient=15.20; 95% CI 5.28-25.11; *P*=.003) between before and after CHEC policy implementation. Meanwhile, STEMI cases demonstrated a significant increase in diagnostic fees (interaction β coefficient=1983.75; 95% CI 84.28-3883.21; *P*=.04) and a significant decrease in upward transfer rate (interaction β coefficient=−0.59; 95% CI –1.18 to –0.001; *P*=.049) than the major trauma cases. Moreover, there were trends toward increasing major treatment indicators (interaction β coefficient=0.30; 95% CI –0.03 to 0.62; *P*=.07), medical orders (interaction β coefficient=11.92; 95% CI –0.90 to 24.73; *P*=.07), and medical expense (interaction β coefficient=24,275.54; 95% CI –640.71 to 4,991,991.78; *P*=.06), although these were not statistically significant. Compared to major trauma cases, no significant change was observed in either process or outcome quality indicators for septic shock cases.

Discussion

Principal Findings

The CHEC policy was universally implemented across Taiwan's emergency medical service systems. Evaluating the impact of such a policy on a nationwide population presents significant challenges, primarily due to the absence of a control group. This limitation restricts the analysis to pre- and postimplementation comparisons, complicating the understanding of how the policy may influence shifts in various diseases. Despite these challenges, our study aimed to analyze the policy's differential effects on time-sensitive conditions thoroughly. By examining disease-specific guidelines, we identified the groups most affected by the policy and those serving as relatively unaffected counterfactuals, offering valuable insights into the policy's impact. To address these challenges, we used a DID design combined with PSM, using major trauma cases as a reference group. This approach ensured that baseline covariates were balanced between before and after policy periods, allowing us to investigate the policy's effects while controlling for preexisting differences in participant and hospital characteristics. Our analysis revealed 2 major findings. First, the CHEC policy effectively improved system efficiency and patient outcomes, resulting in significant reductions in medical orders (ie, –7.29 items per case), short- and long-term mortality rates (ie, –0.09% each), and total medical expenses (ie, –5328.35 points per case), despite a modest increase in diagnostic fees (ie, 376.37 points). Second, we observed unintended "policy spotlight effects," where conditions with time-based surveillance indicators, such as AIS and STEMI, showed disproportionately greater improvements than conditions without such indicators. Specifically, AIS cases experienced significant increases in major treatments (β=0.77) and medical orders (β=15.20), while STEMI cases demonstrated increased diagnostic fees (β=1983.75) and decreased upward transfer rates (β=–0.59) relative to major trauma cases. These findings highlight the varied impacts of the policy based on the presence or absence of time-based monitoring indicators. Thus, the effectiveness and efficiency of the CHEC policy underscore its dual ability to lower costs while improving patient outcomes. However, why has the CHEC policy significantly reduced medical costs and mortality rates, even though the major diagnosis indicator has declined, and no substantial changes have been observed in treatment indicators? The subsequent sections will delve further into diseases' individual and interactive effects to provide analysis.

Distinction Between the Hawthorne and Policy Spotlight Effects

To distinguish health care providers' behaviors influenced by the Hawthorne effect or policy spotlight effects, we selected AIS and STEMI, which have well-established guidelines and

time-based quality indicators under the CHEC policy. In contrast, septic shock, a disease with well-established guidelines but no specific time-based quality indicators, and major trauma, lacking both well-established guidelines and time-based quality indicators, were chosen as reference groups (Table S2 in Multimedia Appendix 1).

After implementing the CHEC policy, we hypothesized that the observed responses might indicate varying levels of awareness among emergency care providers [24]. The Hawthorne effect highlights that individuals' productivity in experimental settings may increase due to the simple fact that they are being observed. This phenomenon underscores the influence of human attention and intervention on behavior [25]. The policy spotlight effects may be intensified by factors such as time constraints, ambiguous symptom patterns, and time-based quality surveillance indicators, prompting emergency care providers to unconsciously adopt selective behaviors concentrating on specific diseases according to policy targets [25]. Consequently, diseases not prioritized by the policy, such as septic shock and major trauma in this study, may experience a significant decline in process quality and a reduction in medical orders on an individual disease basis (Table 3). When using major trauma as the reference group, both AIS and STEMI group showed increases in major treatments, medical orders, diagnostic fees, and medical expenses. In contrast, no significant changes were observed in the septic shock group (Table 4).

When examining the effects of the policy on individual diseases, we observed distinct patterns. A decrease in diagnostic indicators was noted for AIS cases, likely due to increased upward transfer rates, as hospital physicians appeared more inclined to transfer patients with AIS to higher-level hospitals to enhance access to thrombolytic treatment [26]. This aligns with a significant increase in primary treatment indicators. For STEMI cases, there was an increase in diagnostic fees and a notable rise in diagnostic and treatment indicators, coupled with decreased transfer rates, suggesting a more intensive treatment approach that correlates with reduced mortality rates. These findings highlight the differential impacts of the policy, with AIS and STEMI cases benefiting from improvements in process and treatment quality due to the presence of time-based indicators and targeted monitoring. In contrast, septic shock and major trauma cases, which lack such indicators, showed either no significant changes or potential deterioration in quality metrics. These contrasting trends between diseases with and without time-based monitoring suggest the influence of policy spotlight effects. Under the CHEC policy, AIS and STEMI cases have demonstrated marked improvements, while septic shock and major trauma cases remain relatively unaffected, underscoring the importance of targeted policy monitoring in driving quality improvements.

While the Hawthorne effect typically manifests as generalized behavioral improvements across most observed conditions due to the awareness of being monitored, the implementation of the CHEC policy was explicitly designed to optimize emergency medical capacity through enhanced adherence to evidence-based guidelines, with a particular emphasis on reducing unnecessary medical orders and improving emergency care efficiency. The observed outcomes demonstrated significant improvements, including reductions in medical orders and per-case expenses, concurrent with decreased short- and long-term mortality rates, which are the overall impact of CHEC policy effects on the 4 CTSDs. These observed improvements may be attributed to the Hawthorne effect, wherein the awareness of being observed leads to enhanced performance among health care providers.

Both AIS and STEMI cases appeared to be influenced by the policy spotlight effect, and they showed opposite results in diagnostic indicators and upward transfer rates. These could potentially be attributed to the inherent differences between AIS and STEMI cases. For instance, only about 1% to 2% of all patients with AIS underwent primary treatment with intravenous thrombolytic agents, which is far lower than the STEMI major treatment rate. Unexpectedly, in less policy spotlight affected groups, such as septic shock and major trauma, despite significant decreases in primary diagnostic indicators, medical orders, and diagnostic fees, there were unexpected reductions in 30-day and 1-year mortality rates and medical expenses per event. These unintended consequences might resonate with the "less is more" initiative [27], suggesting that curbing excessive diagnoses and avoiding unnecessary procedures could improve patient outcomes and decrease costs [28,29].

Policy Implications

Many emergency care policies implement time-based criteria [30-32], such as Australia's 4-hour rule [30] and the United Kingdom's 4-hour standard [32]. Australia's experience showed that an emergency care policy using time-based criteria could improve emergency congestion without increasing the rate of ED revisits. However, in New Zealand, a policy in effect from 2006 to 2012 dictated that emergency patients must be hospitalized, transferred, or discharged within 6 hours of visiting the ED. After emergency care policy intervention, the length of ED stay decreased while the treatment outcomes of acute myocardial infarction, severe septic shock, and acute appendicitis did not improve significantly [31]. Similarly, after the Canadian Emergency Observation Reduction Program implementation, the length of ED stay decreased, while the treatment quality indicators for acute myocardial infarction, asthma, and upper limb fractures could only be treated in time for the abovementioned time-sensitive diseases during the noncongested emergency period [33].

Emergency care quality is closely related to the practices of medical care providers [12]. Policy makers must reconsider the conventional emphasis on time-based process indicators. This approach may have unintended consequences for diseases that are outside the "spotlight" of rigid time-sensitive evaluation. Instead, a broader and more nuanced evaluation of emergency care quality is needed to incorporate the complexity and ambiguity of various time-sensitive diseases that emergency care providers often manage [12]. We propose replacing time-based indicators with a broader set of multidimensional performance-based indicators, such as those exemplified in the National Health Service Best Practice Tariff policy [34]. This shift toward "best practice" can create a more flexible approach that goes beyond merely relying on time-based or diagnosis-based practices [34]. Such a transition is crucial, as an excessive reliance on diagnostic tests may lead to ED

crowding, a decline in emergency care quality, and increased safety issues [28].

Strengths

This study used a robust DID method within a GEE framework to evaluate the impact of the CHEC policy. The analysis compared pre- and postintervention periods and distinguished treatment and control groups using a categorical disease variable. Interaction terms captured differential policy effects across disease groups, while GEE accounted for correlated data from repeated measures, ensuring robust variance estimation. The estimated coefficients quantified both overall and disease-specific policy impacts, providing a reliable and nuanced evaluation of the CHEC policy. Our study also examined the behaviors of emergency care providers under time constraints and their interactions with strict time-based quality surveillance indicators and adherence to "get with the guidelines" protocols, revealing policy spotlight effects.

Limitations

This study has several limitations. First, the data were derived from a secondary dataset of insurance claims rather than a randomized controlled trial, which may limit causal inferences. Second, the analysis lacked detailed information on time-related quality indicators, such as door-to-evaluation and door-to-treatment times, and critical emergency care metrics. Third, the effectiveness of emergency medical care often depends on collaboration between emergency and consulting physicians; however, this study did not explicitly evaluate the dynamics of their interaction. Future research should examine the coordination and teamwork between these physician groups to understand their impact on care quality better.

Conclusions

Our study reveals that CHEC policy implementation demonstrates a dual capability to reduce costs and improve patient outcomes. The policy spotlight effects result in a disproportional improvement in disease guideline adherence and process quality of CTSDs with time-based surveillance indicators. In contrast, disease entities not fully encompassed in the surveillance indicators may be jeopardized by decreasing diagnosis and treatment process quality.

Acknowledgments

The authors would like to acknowledge the Taiwan National Health Insurance Research Database, provided by the National Health Insurance Administration and managed by the National Health Research Institutes. They would also like to thank the Center for Public Health, Department of Education and Research, Taipei City Hospital, for providing administrative and research grant support. The authors extend their deepest gratitude to the esteemed members of the dissertation committee for their invaluable guidance and support throughout their research journey. The concept of this paper, titled "The Structure, Process, and Outcome Quality Impacts of the Categorization of Hospital Emergency Capability Policy," was significantly shaped by their insightful feedback and thoughtful encouragement, for which they are profoundly grateful.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Generative AI tools (eg, ChatGPT by OpenAI) were used to assist with language refinement and reformatting text for clarity. All AI-generated content was thoroughly reviewed, validated, and edited by the authors to ensure accuracy, scientific rigor, and adherence to academic standards. No content was directly incorporated without author verification, ensuring the integrity of the manuscript remains uncompromised.

Data Availability

The data that support the findings of this study are available from the Taiwan National Health Insurance Research Database, but restrictions apply to the availability of these data, which were used under license for this study, and so are not publicly available. Data are, however, available at the academic request and with the permission of the Taiwan National Health Insurance Administration.

Authors' Contributions

CYL took the lead in conceptualizing the study, drafting the original manuscript, and conducting the formal data analysis. In addition, CYL verified the underlying data presented in the paper. CCL contributed to the study design, data curation, and formal data analysis and oversaw the data collection process. Both CCL and YTH ensured the accuracy of the data analysis and interpretation, as well as verified the paper's underlying data. YCL supervised the study, validated the findings, and played a significant role in reviewing and editing the manuscript. All authors contributed to the development of the study concept and design, as well as the analysis and interpretation of data and the preparation of the manuscript. All authors have approved the final version of the manuscript and take collective responsibility for all aspects of the work, pledging to investigate and resolve any concerns related to the accuracy or integrity of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Propensity score–matched patient characteristics for 4 time-sensitive diseases before and after categorization of hospital emergency capability (CHEC) policy implementation, association of CHEC policy with process and outcome quality in 4 critical time-sensitive diseases, and the distinction between the Hawthorne and policy spotlight effects.
[DOC File , 157 KB - ijmr_v14i1e54651_app1.doc]

Multimedia Appendix 2
Association of CHEC policy with process and outcomes quality in four critical time-sensitive diseases.
[DOCX File , 25 KB - ijmr_v14i1e54651_app2.docx]

References

1. AHRQ quality indicators—guide to prevention quality indicators: hospital admission for ambulatory care sensitive conditions. Agency for Healthcare Research and Quality. 2001. URL: https://www.ahrq.gov/downloads/pub/ahrqqi/pqiguide.pdf [accessed 2025-02-13]
2. Combes JR, Arespacochaga E. Appropriate use of medical resources. American Hospital Association's Physician Leadership Forum. 2013. URL: https://www.aha.org/system/files/content/13/appropusewhiteppr.pdf [accessed 2025-02-13]
3. Panchavati S, Lam C, Zelin NS, Pellegrini E, Barnes G, Hoffman J, et al. Retrospective validation of a machine learning clinical decision support tool for myocardial infarction risk stratification. Healthc Technol Lett 2021 Dec 31;8(6):139-147 [FREE Full text] [doi: 10.1049/htl2.12017] [Medline: 34938570]
4. Daley S, Braimah J, Sailor S, Kongable GL, Barch C, Rapp K, et al. Education to improve stroke awareness and emergent response. The NINDS rt-PA Stroke Study Group. J Neurosci Nurs 1997 Dec;29(6):393-396. [doi: 10.1097/01376517-199712000-00009] [Medline: 9479661]
5. Fesmire FM, Brady WJ, Hahn S, Decker WW, Diercks DB, Ghaemmaghami CA, et al. Clinical policy: indications for reperfusion therapy in emergency department patients with suspected acute myocardial infarction. American College of Emergency Physicians Clinical Policies Subcommittee (Writing Committee) on reperfusion therapy in emergency department patients with suspected acute myocardial infarction. Ann Emerg Med 2006 Oct;48(4):358-383. [doi: 10.1016/j.annemergmed.2006.07.006] [Medline: 16997672]
6. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001 Nov 08;345(19):1368-1377. [doi: 10.1056/NEJMoa010307] [Medline: 11794169]
7. Carr BG, Kilaru AS, Karp DN, Delgado MK, Wiebe DJ. Quality through coopetition: an empiric approach to measure population outcomes for emergency care-sensitive conditions. Ann Emerg Med 2018 Sep;72(3):237-245 [FREE Full text] [doi: 10.1016/j.annemergmed.2018.03.004] [Medline: 29685369]
8. Mehrotra A, Sklar DP, Tayal VS, Kocher KE, Handel DA, Myles Riner R. Important historical efforts at emergency department categorization in the United States and implications for regionalization. Acad Emerg Med 2010 Dec;17(12):e154-e160 [FREE Full text] [doi: 10.1111/j.1553-2712.2010.00931.x] [Medline: 21122008]
9. Haupt MT, Bekes CE, Brilli RJ, Carl LC, Gray AW, Jastremski MS, et al. Guidelines on critical care services and personnel: recommendations based on a system of categorization of three levels of care. Crit Care Med 2003 Nov;31(11):2677-2683. [doi: 10.1097/01.CCM.0000094227.89800.93] [Medline: 14605541]
10. Glickman SW, Lytle BL, Ou FS, Mears G, O'Brien S, Cairns CB, et al. Care processes associated with quicker door-in-door-out times for patients with ST-elevation-myocardial infarction requiring transfer: results from a statewide regionalization program. Circ Cardiovasc Qual Outcomes 2011 Jul;4(4):382-388. [doi: 10.1161/CIRCOUTCOMES.110.959643] [Medline: 21712523]
11. Govindarajan P, Shiboski S, Grimes B, Cook LJ, Ghilarducci D, Meng T, et al. Effect of acute stroke care regionalization on intravenous alteplase use in two urban counties. Prehosp Emerg Care 2020 Nov 11;24(4):505-514. [doi: 10.1080/10903127.2019.1679303] [Medline: 31599705]
12. Rubenstein LV, Mittman BS, Yano EM, Mulrow CD. From understanding health care provider behavior to improving health care: the QUERI framework for quality improvement. Med Care 2000;38(6):I-129-I-141. [doi: 10.1097/00005650-200006001-00013]
13. Gilovich T, Savitsky K. The spotlight effect and the illusion of transparency: egocentric assessments of how we are seen by others. Curr Dir Psychol Sci 1999 Dec 01;8(6):165-168. [doi: 10.1111/1467-8721.00039]
14. Salisbury KM. National and state policies influencing the care of children affected by AIDS. Child Adolesc Psychiatr Clin N Am 2000 Apr;9(2):425-449. [doi: 10.1016/s1056-4993(18)30128-7]
15. Savitsky K, Epley N, Gilovich T. Do others judge us as harshly as we think? Overestimating the impact of our failures, shortcomings, and mishaps. J Pers Soc Psychol 2001 Jul;81(1):44-56. [doi: 10.1037//0022-3514.81.1.44] [Medline: 11474724]
16. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the National Health Insurance Research Database (NHIRD). Epidemiol Health 2018;40:e2018062 [FREE Full text] [doi: 10.4178/epih.e2018062] [Medline: 30727703]

17. Lin CY, Lee YC. Effectiveness of hospital emergency department regionalization and categorization policy on appropriate patient emergency care use: a nationwide observational study in Taiwan. BMC Health Serv Res 2021 Jan 06;21(1):21 [FREE Full text] [doi: 10.1186/s12913-020-06006-7] [Medline: 33407444]
18. Hospital Emergency Capability Categorization. Ministry of Health and Welfare Taiwan. URL: https://dep.mohw.gov.tw/DOMA/np-980-106.html [accessed 2025-02-27]
19. Shafi S, Nathens AB, Cryer HG, Hemmila MR, Pasquale MD, Clark DE, et al. The trauma quality improvement program of the American College of Surgeons Committee on Trauma. J Am Coll Surg 2009 Oct;209(4):521-30.e1. [doi: 10.1016/j.jamcollsurg.2009.07.001] [Medline: 19801325]
20. Cox S, Smith K, Currell A, Harriss L, Barger B, Cameron P. Differentiation of confirmed major trauma patients and potential major trauma patients using pre-hospital trauma triage criteria. Injury 2011 Sep;42(9):889-895. [doi: 10.1016/j.injury.2010.03.035] [Medline: 20430387]
21. Dobaria V, Kwon OJ, Hadaya J, Sanaiha Y, Sareh S, Aguayo E, et al. Impact of center volume on outcomes of surgical repair for type A acute aortic dissections. Surgery 2020 Jul;168(1):185-192. [doi: 10.1016/j.surg.2020.04.007] [Medline: 32507629]
22. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CAJ. A conceptual model of emergency department crowding. Ann Emerg Med 2003 Aug;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
23. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May 31;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
24. Lied TR, Kazandjian VA. A Hawthorne strategy: implications for performance measurement and improvement. Clin Perform Qual Health Care 1998;6(4):201-204. [Medline: 10351289]
25. Franke RH, Kaul JD. The Hawthorne experiments: first statistical interpretation. Am Sociol Rev 1978 Oct;43(5):623-643. [doi: 10.2307/2094540]
26. Zachrison KS, Amati V, Schwamm LH, Yan Z, Nielsen V, Christie A, et al. Influence of hospital characteristics on hospital transfer destinations for patients with stroke. Circ Cardiovasc Qual Outcomes 2022 May;15(5):e008269. [doi: 10.1161/CIRCOUTCOMES.121.008269] [Medline: 35369714]
27. Wegwarth O, Gigerenzer G. Less is more: overdiagnosis and overtreatment: evaluation of what physicians tell their patients about screening harms. JAMA Intern Med 2013 Dec 09;173(22):2086-2087. [doi: 10.1001/jamainternmed.2013.10363] [Medline: 24145597]
28. Kawano T, Nishiyama K, Hayashi H. Execution of diagnostic testing has a stronger effect on emergency department crowding than other common factors: a cross-sectional study. PLoS One 2014 Oct 13;9(10):e108447 [FREE Full text] [doi: 10.1371/journal.pone.0108447] [Medline: 25310089]
29. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. Ann Emerg Med 2008 Aug;52(2):126-136 [FREE Full text] [doi: 10.1016/j.annemergmed.2008.03.014] [Medline: 18433933]
30. Ngo H, Forero R, Mountain D, Fatovich D, Man WN, Sprivulis P, et al. Impact of the four-hour rule in Western Australian hospitals: trend analysis of a large record linkage study 2002-2013. PLoS One 2018 Mar 14;13(3):e0193902 [FREE Full text] [doi: 10.1371/journal.pone.0193902] [Medline: 29538401]
31. Jones P, Le Fevre J, Harper A, Wells S, Stewart J, Curtis E, et al. Effect of the shorter stays in emergency departments time target policy on key indicators of quality of care. N Z Med J 2017 May 12;130(1455):35-44. [Medline: 28494476]
32. Cooke MW. Reforming the UK emergency care system. Emerg Med J 2003 Mar;20(2):113-114 [FREE Full text] [doi: 10.1136/emj.20.2.113] [Medline: 12642515]
33. Vermeulen MJ, Guttmann A, Stukel TA, Kachra A, Sivilotti ML, Rowe BH, et al. Are reductions in emergency department length of stay associated with improvements in quality of care? A difference-in-differences analysis. BMJ Qual Saf 2016 Jul 13;25(7):489-498 [FREE Full text] [doi: 10.1136/bmjqs-2015-004189] [Medline: 26271919]
34. Copas DP, Moran CG. Major trauma care in England: changing the state of a nation's healthcare system. Bone Joint 360 2014 Apr;3(2):2-5. [doi: 10.1302/2048-0105.32.360217]

Abbreviations

AIS: acute ischemic stroke
CHEC: categorization of hospital emergency capability
CTSD: critical time-sensitive disease
DID: difference-in-differences
ED: emergency department
GEE: generalized estimating equation
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
LHID2005: Longitudinal Health Insurance Database
PSM: propensity score matching
STEMI: ST-segment elevation myocardial infarction

Edited by T de Azevedo Cardoso; submitted 26.11.23; peer-reviewed by A Debbarma, T Cinar, M Al Zoubi; comments to author 06.01.25; revised version received 20.01.25; accepted 23.01.25; published 25.03.25.

Please cite as:
Lin CY, Liu CC, Huang YT, Lee YC
Policy Spotlight Effects on Critical Time-Sensitive Diseases: Nationwide Retrospective Cohort Study on Taiwan's Hospital Emergency Capability Categorization Policy
Interact J Med Res 2025;14:e54651
URL: https://www.i-jmr.org/2025/1/e54651
doi:10.2196/54651
PMID:

Original Paper

Current Status and Future Directions of Ferroptosis Research in Breast Cancer: Bibliometric Analysis

Jia-Yuan Luo[1*], MD; Yu-Long Deng[2*], BMed; Shang-Yi Lu[3], BMed; Si-Yan Chen[4], BMed; Rong-Quan He[2], MD, PhD; Di-Yuan Qin[5], BS; Bang-Teng Chi[2], BMed; Gang Chen[1], MD, PhD; Xia Yang[1], MD, PhD; Wei Peng[2], MD, PhD

[1]Department of Pathology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[2]Department of Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[3]Department of Hepatological and Gland Surgery, Wuzhou Gongren Hospital/The Seventh Affiliated Hospital of Guangxi Medical University, Wuzhou, China

[4]Day Chemotherapy Center, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[5]Department of Computer Science and Technology, School of Computer and Electronic Information, Guangxi University, Nanning, China

[*]these authors contributed equally

Corresponding Author:
Wei Peng, MD, PhD
Department of Oncology, The First Affiliated Hospital of Guangxi Medical University
No.6 Shuangyong Road
Nanning, 530021
China
Phone: 86 15977779804
Email: pengwei@stu.gxmu.edu.cn

Abstract

Background: Ferroptosis, as a novel modality of cell death, holds significant potential in elucidating the pathogenesis and advancing therapeutic strategies for breast cancer.

Objective: This study aims to comprehensively analyze current ferroptosis research and future trends, guiding breast cancer research advancements and innovative treatment strategies.

Methods: This research used the R package Bibliometrix (Department of Economic and Statistical Sciences at the University of Naples Federico II), VOSviewer (Centre for Science and Technology Studies at Leiden University), and CiteSpace (Drexel University's College of Information Science and Technology), to conduct a bibliometric analysis of 387 papers on breast cancer and ferroptosis from the Web of Science Core Collection. The analysis covers authors, institutions, journals, countries or regions, publication volumes, citations, and keywords.

Results: The number of publications related to this field has surged annually, with China and the United States collaborating closely and leading in output. Sun Yat-sen University stands out among the institutions, while the journal *Frontiers in Oncology* and the author Efferth T contribute significantly to the field. Highly cited papers within the domain primarily focus on the induction of ferroptosis, protein regulation, and comparisons with other modes of cell death, providing a foundation for breast cancer treatment. Keyword analysis highlights the maturity of glutathione peroxidase 4-related research, with breast cancer subtypes emerging as motor themes and the tumor microenvironment, immunotherapy, and prognostic models identified as basic themes. Furthermore, the application of nanoparticles serves as an additional complement to the basic themes.

Conclusions: The current research status in the field of ferroptosis and breast cancer primarily focuses on the exploration of relevant theoretical mechanisms, whereas future trends and mechanisms emphasize the investigation of therapeutic strategies, particularly the clinical application of immunotherapy related to the tumor microenvironment. Nanotherapy has demonstrated significant clinical potential in this domain. Future research directions should deepen the exploration in this field and accelerate the clinical translation of research findings to provide new insights and directions for the innovation and development of breast cancer treatment strategies.

(Interact J Med Res 2025;14:e66286) doi:10.2196/66286

KEYWORDS

breast cancer; ferroptosis; bibliometric; malignancy; cancer studies; treatment; bibliometric analysis; VOSviewer; China; United States; breast carcinoma; mammary cancer; strategy; trends; bibliography; review; disparities; forecast; treatment strategies; advancements

Introduction

Based on 2022 International Agency for Research on Cancer data, the global burden of breast cancer is alarming, with 2.3 million new cases, and over 665,000 deaths annually. Breast cancer constitutes 11.6% of new cancers and 6.9% of cancer deaths, underscoring the need for urgent research and intervention [1]. It has been projected that by 2040, due to global population growth and aging, the cancer burden will further increase, with breast cancer experiencing particularly notable growth. The number of new cases and deaths are expected to rise to over three million and one million, respectively [2]. As the most prevalent cancer worldwide, surpassing lung cancer, breast cancer poses a significant threat to human health. Despite the considerable progress achieved through sustained research in breast cancer diagnosis, treatment, and prevention, the intricate etiology of this malignancy remains widely undeciphered. Consequently, there is an urgent need to delve deeper into the complex mechanisms that underlie its development, with the aim of developing more precise and efficacious strategies for its management [3]. There is considerable variation in treatment response among individuals with this disease, and the prognosis and treatment outcomes for individuals diagnosed with advanced breast cancer are notably unfavorable [4]. Exploring new mechanisms can enhance our understanding of the biological behavior of breast cancer and improve treatment sensitivity.

Ferroptosis has added a new theoretical basis to the study of the molecular mechanisms underlying breast cancer. Since the introduction of this concept, the popularity of research in this field has rapidly increased, showing an exponential growth trend [5]. Ferroptosis is a recently discovered cell death mechanism mediated by iron-dependent phospholipid peroxidation and is tightly regulated by various cellular metabolic networks, including redox balance, iron handling, mitochondrial function, and amino acid, lipid, and glucose metabolism. This intricate regulation highlights its significance in breast cancer research [6]. Studies on breast cancer have revealed that the regulatory network of ferroptosis is sophisticated and complex, involving the synergistic effects of multiple signaling pathways and transcription factors. This regulatory network may modulate the malignant biological behavior of breast cancer cells via autophagy [7], metabolism [8], immune regulation [9], and other procedures, opening novel strategic directions for breast cancer treatment [5,6,10]. However, the literature concerning breast cancer and ferroptosis currently appears highly fragmented, and a systematic bibliometric analysis to consolidate this information is lacking, making it difficult to accurately summarize the research status and future direction of this field.

In response, this study aims to use bibliometric methods to conduct a comprehensive and objective analysis of the current status and future trends in breast cancer and ferroptosis research. This will entail a rigorous examination of pertinent indicators, including author distribution, research institutions, journals, country or region contributions, scientific publication output, highly cited papers, and keyword analysis. The ultimate objective of this study is to provide an in-depth revelation of the research hotspots and potential development trends in this field, thereby offering guidance for future research endeavors in exploring the mechanisms and therapeutic strategies related to ferroptosis in breast cancer.

Methods

Data Sources and Search Strategy

This study used data from two core collections within the Web of Science Core Collection (WOSCC): the Social Sciences Citation Index and the Science Citation Index Expanded. An advanced search was conducted using the search query (TS=(“Breast Tumor*” OR “Breast Cancer” OR “Mammary Cancer” OR “Breast Carcinoma” OR “Carcinomas, Breast”) AND TS = (“Ferroptosis”)), yielding a total of 624 papers. The search was further refined to include papers published between January 1, 2016, and February 1, 2024. Paper types were limited to original research papers, excluding reviews, conference abstracts, news items, advertisements, and non-English literature that did not meet the inclusion criteria. All data records in this study consisted of complete records and cited references and were exported in plain text format. To avoid errors and ambiguities due to database updates, all searches and data exports were completed on February 18, 2024. Additionally, all data in this study were obtained from open-source public databases; thus, no ethical issues were involved. The search and screening processes were independently conducted by JYL and YLD. In the case of any disagreements during the screening process, the final decision was reached after a discussion with the corresponding authors WP and XY.

Inclusion and Exclusion Criteria

The inclusion criteria were (1) published journal papers; (2) papers with “breast cancer” and “ferroptosis” in the title, indicating a primary focus on breast cancer and ferroptosis; (3) papers not mentioning “ferroptosis” in the title but including ferroptosis-related research on breast cancer in the abstract; and (4) papers containing actual research on the relationship between breast cancer and ferroptosis within the full text.

The exclusion criteria were (1) nonliterature records, such as reviews, conference reports, briefings, news articles, editorial materials, letters, and advertisements; (2) non-English literature, (3) duplicate or data-deficient papers; and (4) withdrawn papers.

Data Standardization and Visualization Methods in Bibliometrics

To ensure data accuracy and consistency, rigorous data standardization processing was applied to the raw data in this study. First, the study uniformly standardized the names of

authors and institutions. Second, controversies over the attribution of country or region names were carefully resolved to ensure that the geographical attributes of the data were clear and unambiguous. In addition, synonymous keywords with the same meaning were merged, and keywords that were not directly related to the specific research content were deleted.

The study used the following tools for visualization and bibliometric analysis: the R package Bibliometrix (version 4.3.2; Department of Economic and Statistical Sciences at the University of Naples Federico II), VOSviewer (version 1.6.19; Centre for Science and Technology Studies [CWTS] at Leiden University), and CiteSpace (version 6.1.R6; Drexel University's College of Information Science and Technology). The specific content included visual analyses of annual scientific publication output and citation trends, authors, institutions, journals, countries or regions, highly cited papers, and keywords. Notably, in VOSviewer, we applied "full counting" for normalized data processing. For cluster analysis of authors, institutions, and countries or regions, a minimum threshold of three publications and citations was set, and the clustering map of the top 100 keywords was further presented. Additionally, the Walktrap method was used in thematic map analysis for more precise topic delineation, while other analyses adhered to default software thresholds. By using these visualization methods, this study provided researchers in related fields with an intuitive and clear understanding of the research landscape.

Ethical Considerations

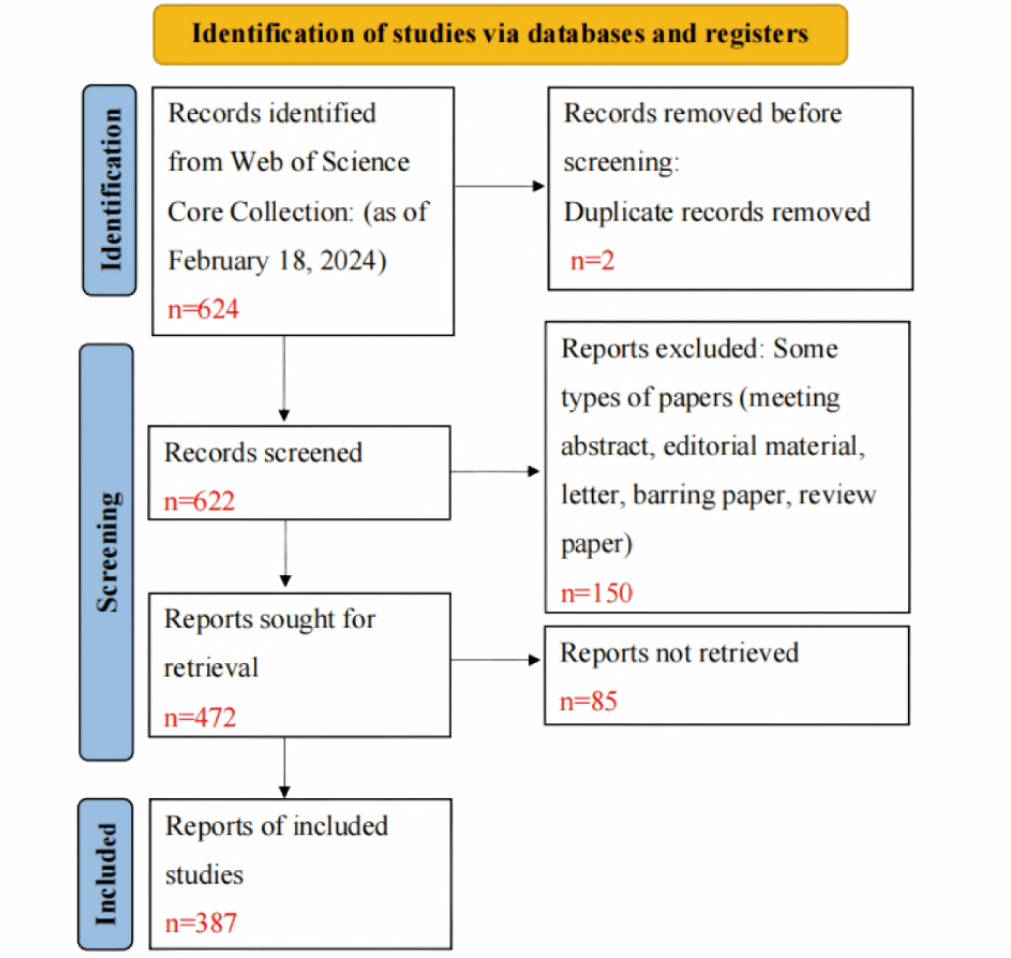
No animal or human studies were carried out by the authors for this paper.

Results

Key Information

This study included a total of 387 academic papers published by 2494 authors in 192 academic journals as the data source (the detailed data collection process is illustrated in Figure 1). On average, each paper had 8.26 coauthors, with only two papers authored by a single individual. The proportion of international collaborations was 16.54%. A total of 840 keywords and 15,107 cited references were identified. It is noteworthy that the average publication age of papers in this field was only 2.02 years, while the average number of citations per paper was as high as 27.22. These data demonstrate that this topic is rapidly emerging as a new research focus (Multimedia Appendix 1).

Figure 1. Flowchart of literature data included in this study.

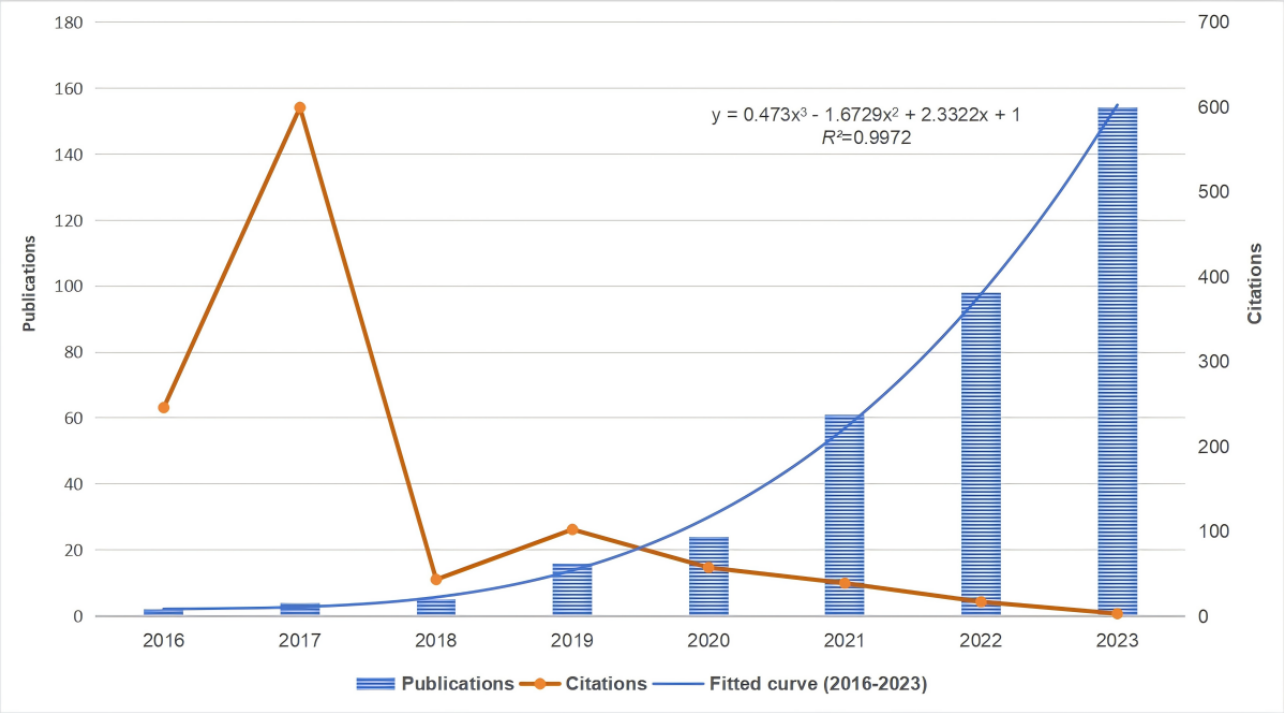


Analysis of Scientific Output in Breast Cancer– and Ferroptosis-Related Research

Annual Analysis of Scientific Publications

Due to the limited availability of relevant literature data published in 2024 up to the retrieval cutoff date of February 1, this study collected complete annual data spanning from 2016 to 2023 (n=364) from a total of 387 papers. A cubic function was fitted to the data to analyze the trend in publication volume (R^2= 0.9972; Figure 2). Specifically, the bar chart indicates that the first paper in this field was published in 2016 (n=1), with 2023 marking the peak year of annual publications (n=154). The line graph represents the average annual citation frequency of relevant literature from 2016 to 2023, revealing a peak of 599 citations in 2017, followed by an overall decreasing trend in subsequent years.

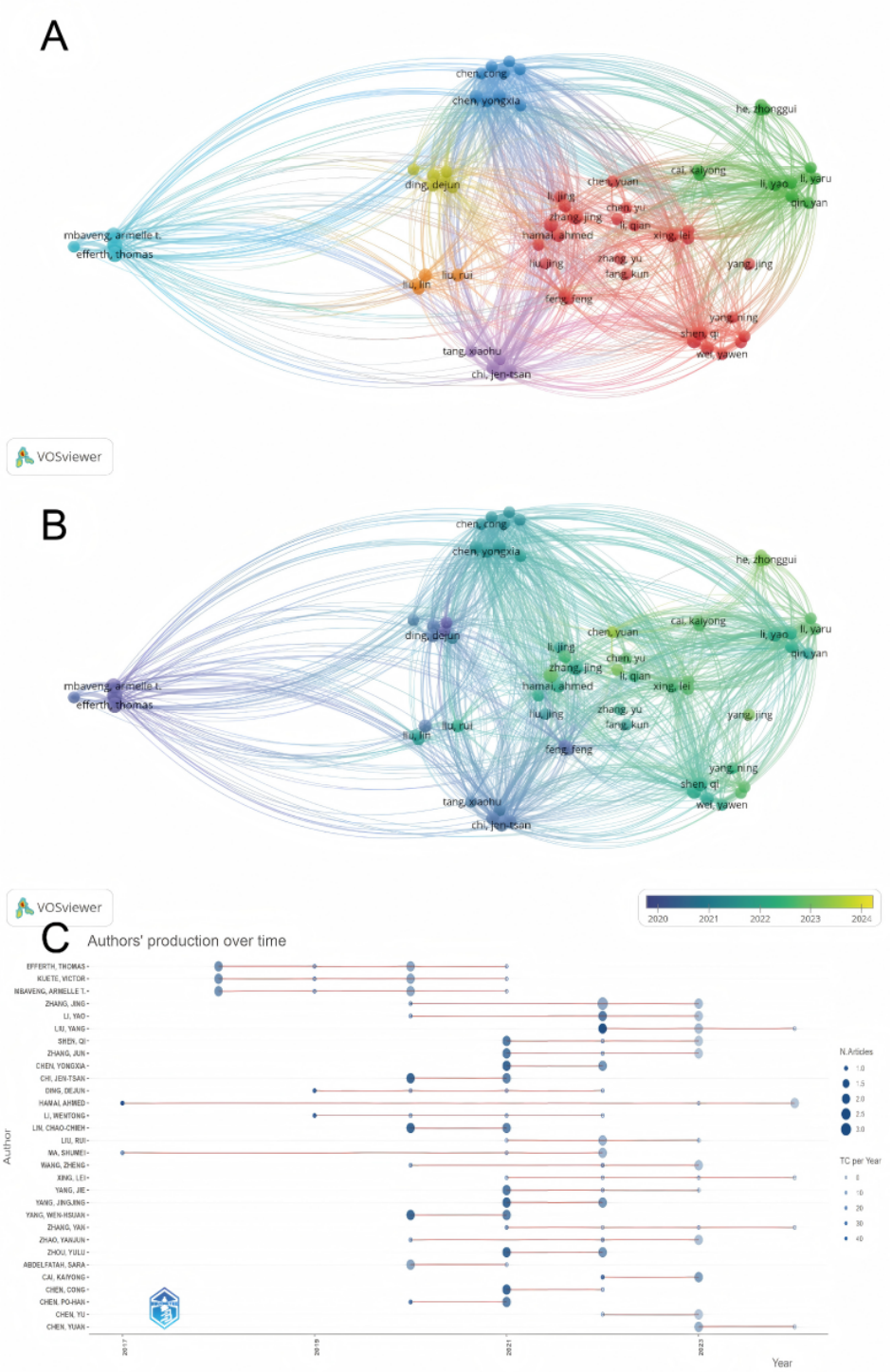
Figure 2. Visualization of annual publication and citation trends of papers (2016-2023). Note that the literature data for 2024 is only counted up to February 1, 2024, with 23 publications and 0 citations.



Analysis of Scientific Publications by Prolific Authors

Through an analysis of the scientific publication output of prolific authors, the study found that 22 distinguished authors had published more than two papers in this research field, with no significant differences in the publication volume among the author groups in the field of ferroptosis and breast cancer. Notably, a minority of authors, represented by Efferth et al, had a significantly higher publication output, reaching six papers. Closely following them were authors from the Shen and Zhang groups, who have also contributed significantly to the lactoferrin research landscape, with a noteworthy publication count of five papers each, occupying important positions in the field of breast cancer and ferroptosis research (Figure 3A). It is noteworthy that, excluding the incomplete data from 2024, as shown in Figures 3B and 3C, an increasing number of new distinguished authors have emerged in recent years, with their publications mainly concentrated after 2020.

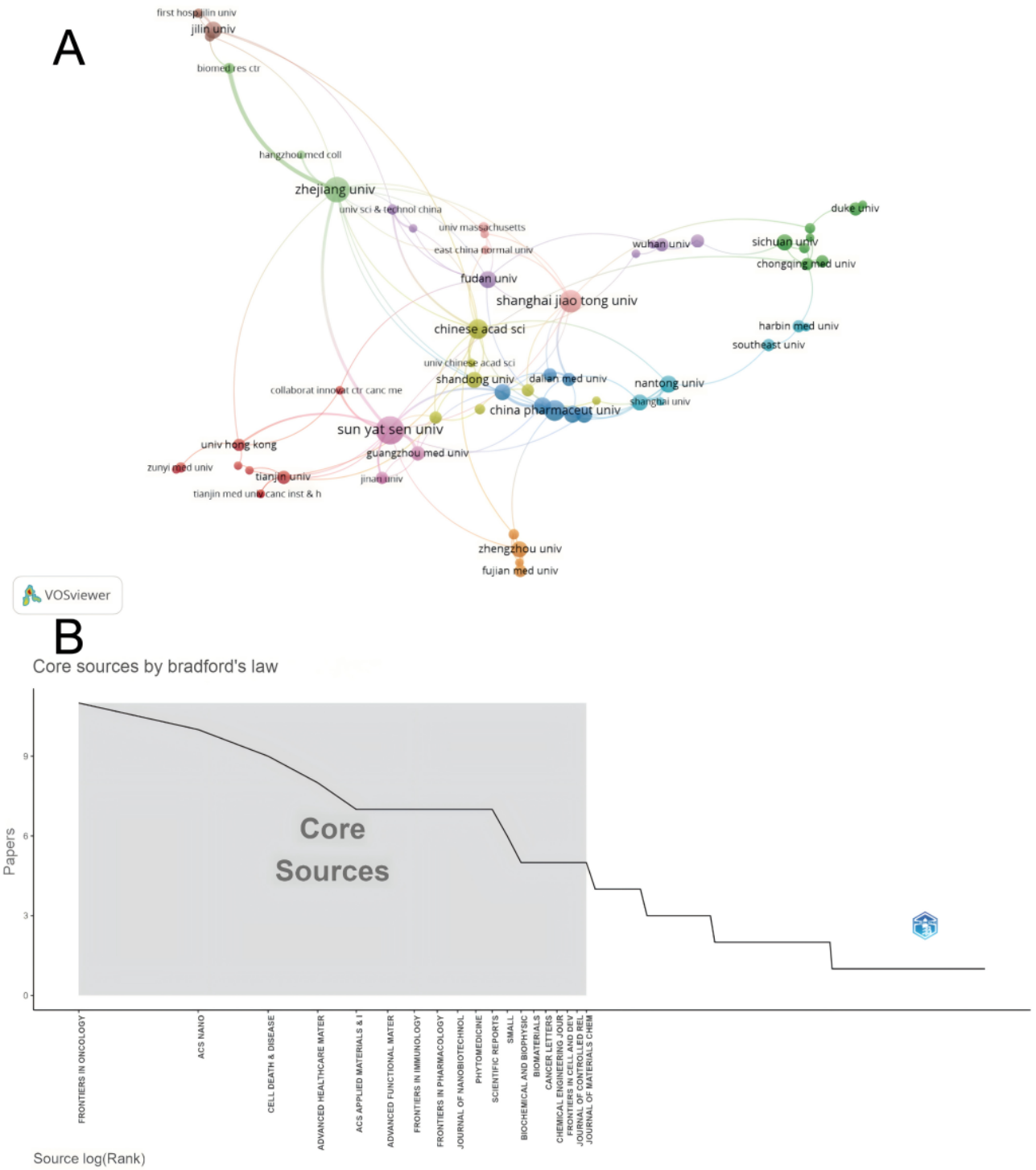
Figure 3. Visualization charts of authors' publication volume and publication trends. (A) Network visualization map of authors' publication volume. (B) Stacked visualization map of authors' publication volume. (C) Heatmap of authors' publication volume.



Analysis of Scientific Publications by Prolific Institutions

By analyzing the bibliometric data related to the research fields of breast cancer and ferroptosis, this study identified a group of institutions with significant scientific output in this area (Figure 4A). Among them, Sun Yat-sen University (n=23 publications) and Zhejiang University (n=19 publications) lead the list of institutions with publications related to breast cancer and ferroptosis. Apart from these two universities, other renowned domestic institutions, such as Shanghai Jiao Tong University (n=15 publications), China Pharmaceutical University (n=13 publications), and the Chinese Academy of Sciences (n=12 publications) have also performed exceptionally well in this field. It is also worth mentioning that non-Chinese institutions, such as Duke University (n=6 publications) and Stanford University (n=3 publications), also possess research capabilities in this area.

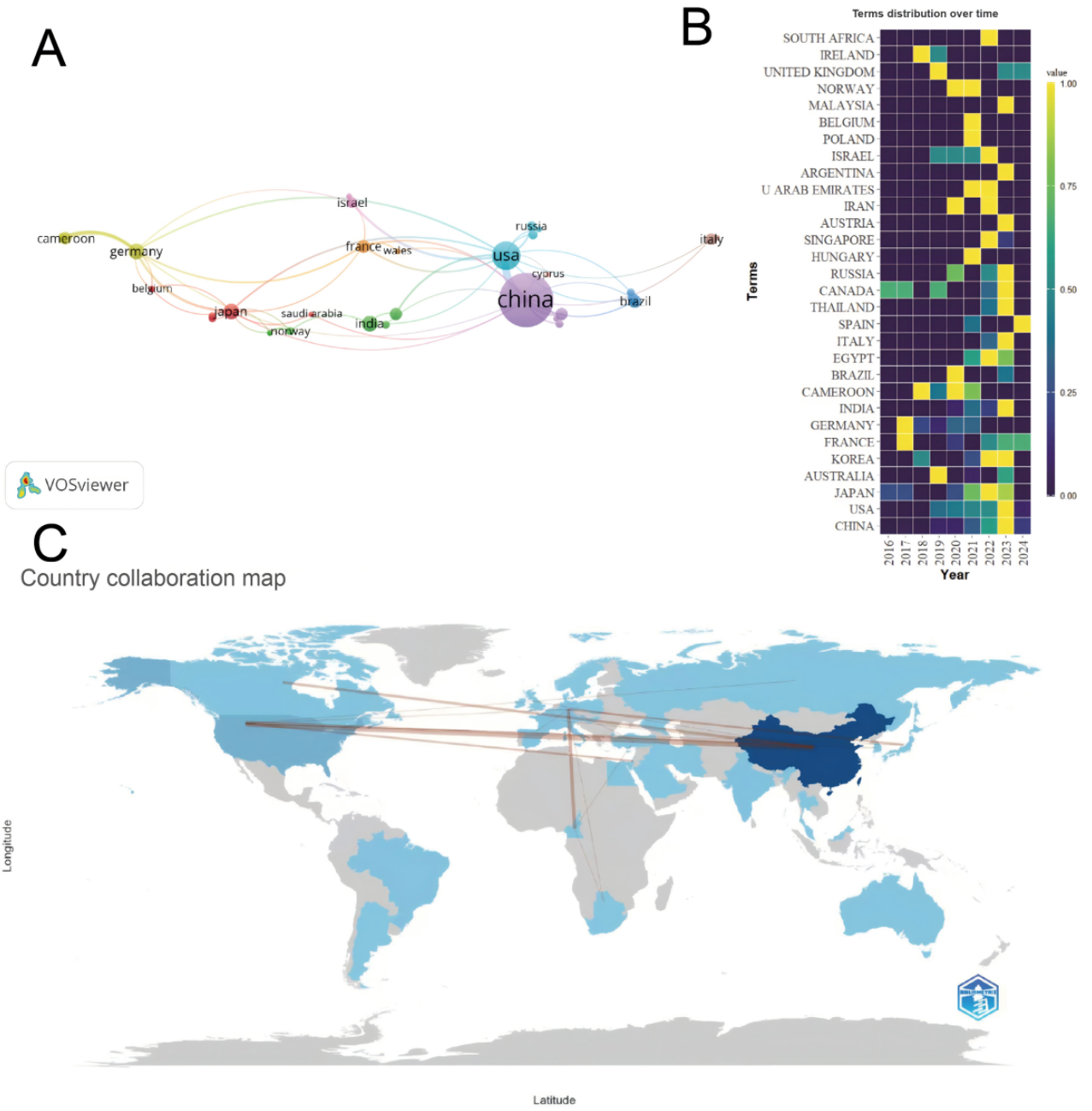
Figure 4. Visualization maps of institutions' publication volume and identification of core journals. (A) Network visualization map of scientific institutions' publication volume. (B) Identification of journals that have made significant contributions to the field, based on Bradford's Law.



Analysis of Scientific Publications by Prolific Countries

The study found that China (n=287) and the United States (n=52) occupy a dominant position in terms of paper output. Furthermore, countries such as Germany (n=11), India (n=11), Japan (n=11), France (n=7), Brazil (n=7), and Cameroon (n=6) have achieved certain research outcomes in the study of breast cancer and ferroptosis (Figure 5A). Furthermore, as depicted in Figure 5B, China experienced a pronounced acceleration in the volume of scientific publications focused on this research area around 2023, exhibiting a remarkable growth pattern. Moreover, the establishment of a close collaborative partnership between China and the United States emphasizes the inherently global nature of research endeavors in this domain (Figure 5C).

Figure 5. Global publication landscape and collaboration in breast cancer and ferroptosis research. (A) Network visualization map of country publication volume. (B) Heatmap of country publication trends. (C) Map of country collaboration relationships.



Analysis of Citation Volume in Breast Cancer– and Ferroptosis-Related Research

Analysis of Frequently Cited Papers

As shown in the data presented in Multimedia Appendix 2, this study found that the highest citation count was from the paper "Ferroptosis is induced following siramesine and lapatinib treatment of breast cancer cells," published in the journal *Cell Death & Disease*, with 38 citations. From the perspective of research focus and content, these citations primarily concentrate on the following aspects: first, the mechanisms and impacts of ferroptosis induction in breast cancer cells by distinctive drugs or therapeutic methods, as exemplified by papers 1, 2, 3, 4, 6, and 10; second, the exploration of the web-based mechanisms between related protein modulation and ferroptosis variations in breast cancer, as demonstrated in papers 2, 5, 7, 8, and 9; and third, the contrasts and interrelationships between ferroptosis and other modes of cell death in breast cancer cells, as discussed in paper 9.

Analysis of Frequently Cited Authors

The contributions and recognition of different authors can be reflected by the total citation count of their papers. According to the author citation analysis in Multimedia Appendix 3, Chen Y ranks at the top among all authors, with a total citation count of 64. Following closely is Gibson SB, with a total citation count of 51. Additionally, papers by Ding DJ, Li WT, Gai CC, Henson EE, Li ZHR, and Ma S have all been cited 38 times or more.

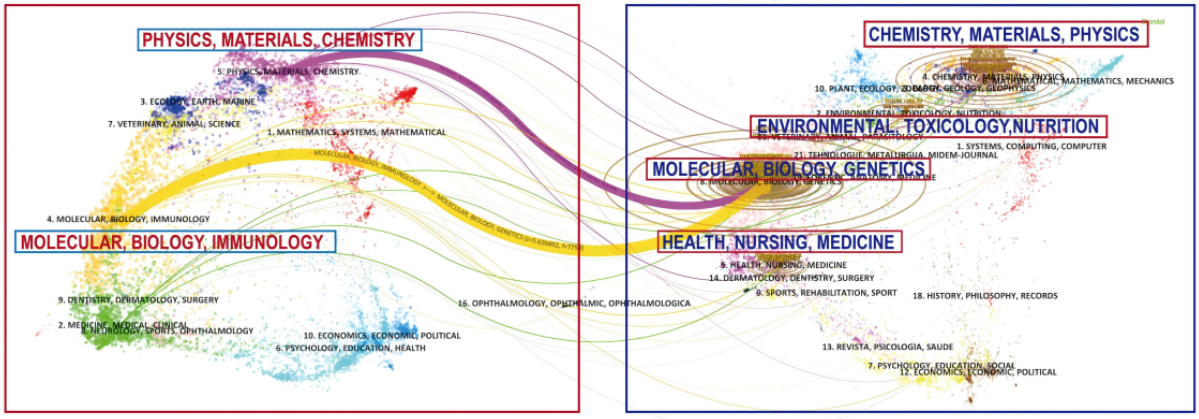
Analysis of Frequently Cited Countries

There are notable differences in the citations of research outcomes from different countries. According to the results shown in Multimedia Appendix 4, China ranks at the top in terms of total citations, with 4877, but the average number of citations per paper is 17.40. Conversely, Germany boasts the highest average citations per paper, reaching an impressive 489.20 citations. The United Kingdom, France, and Canada follow closely with average citations of 187.00, 118.70, and 108.80 per paper, respectively. It is noteworthy that the United States while ranking third after China and Germany in total citations with 1658, records an average of 50.20 citations per paper.

Overlay Analysis of High-Frequency Citing and Cited Journals

In this study, CiteSpace software was used to construct a dual-layer overlay map of journals, visually representing the intricate relationships between citing journals (located on the left) and cited journals (located on the right) within the research domain of breast cancer and ferroptosis (Figure 6). The analysis results revealed that the group of citing journals represented by "Molecular, Biology, Immunology" and the core group of cited journals centered around "Molecular, Biology, Genetics" constitute the pivotal nodes in the knowledge flow path at the journal level (*z* score=5.639492; *f*=7758). Closely following this, there is a strong connection between the citing journal group of "Physics, Materials, Chemistry" and the cited journal group of "Molecular, Biology, Genetics" (*z* score= 2.1142766; *f*=3189).

Figure 6. Dual-map overlay of journals illustrating the citation relationships between citing journals on the left and cited journals on the right. The width of the connecting lines represents the strength of the citation relationships.



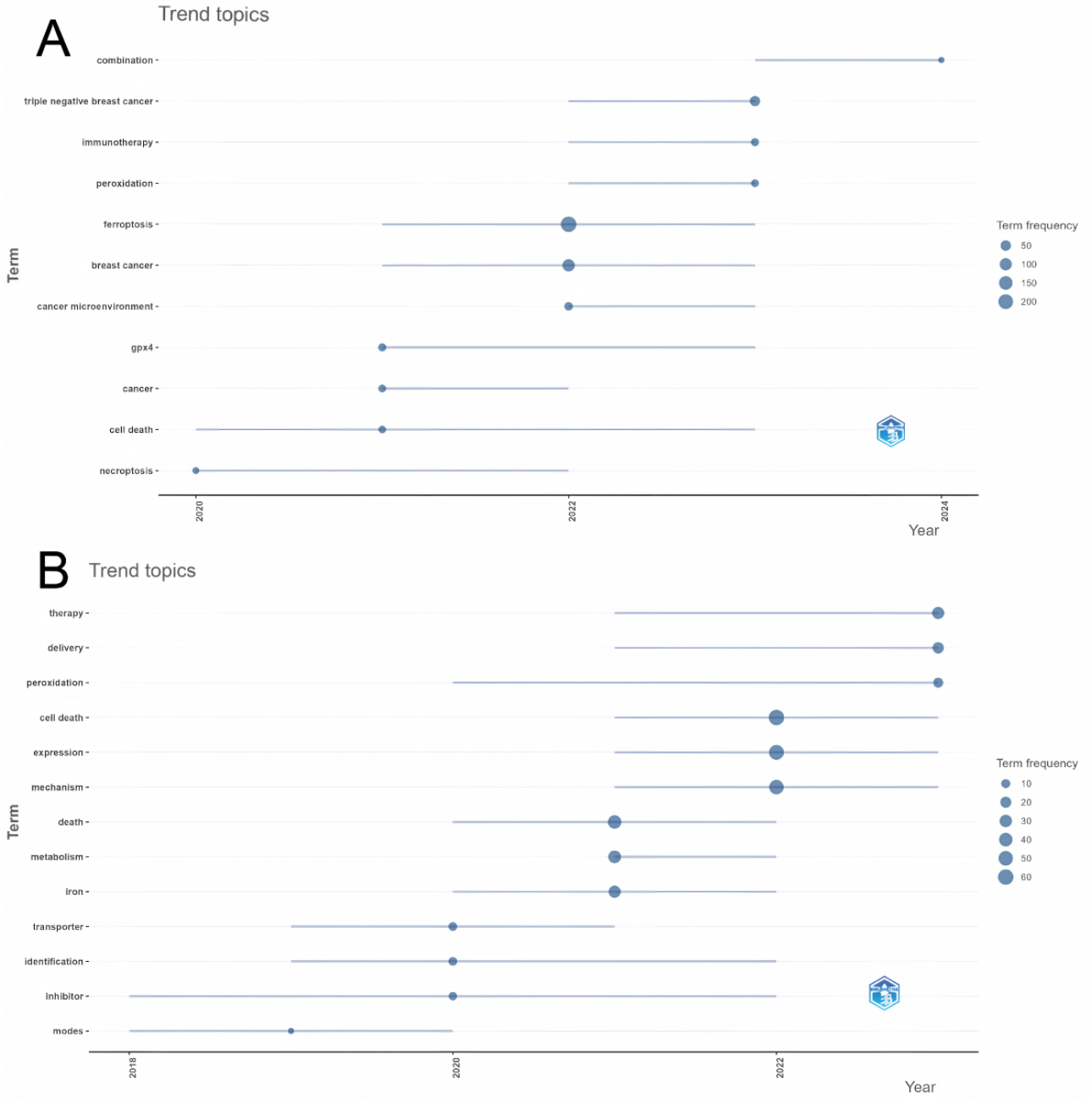
Cluster Analysis of High-Frequency Keywords and Main Evolution Directions in Breast Cancer– and Ferroptosis-Related Research

Cluster Analysis of Author Keywords and Analysis of Main Evolution Trends

After eliminating the core terms "breast cancer" and "ferroptosis," along with their synonyms, this study conducted a clustering analysis of the top 100 most frequent author keywords. The results indicate that these keywords can be categorized into four primary clusters (Multimedia Appendix 5). Specifically, Cluster 1 (in red) predominantly focuses on research topics such as "triple-negative breast cancer" (n=56), "prognosis model" (n=18), and "cancer microenvironment" (n=17). Cluster 2 (in green) centers its research around concepts such as "peroxidation" (n=13), "photothermal therapy" (n=12), and the general term "cancer" (n=11). Cluster 3 (in blue) revolves around high-frequency keywords, such as "reactive oxygen species" (n=17), "gpx4" (n=14), and "stress" (n=13). Finally, Cluster 4 (in orange) concentrates on areas related to "cell death" (n=10), "anticancer" (n=6), and "drug delivery" (n=5).

A deeper analysis was undertaken to examine the evolutionary patterns of high-frequency keywords over recent years, as presented in Multimedia Appendix 6A and Figure 7A. These trends reveal a progressive intensification of research focused on the concepts of cell death and glutathione peroxidase 4 (GPX4), indicating a growing interest and emphasis within the scientific community. It is noteworthy that from 2022 onwards, the frequency of keywords related to "cancer microenvironment," "triple-negative breast cancer," "immunotherapy," "combination therapy," and "peroxidation" increased significantly. Additionally, as witnessed from the time heatmap of the top 30 author keywords, keywords such as "erastin" and "autophagy" have maintained stable popularity in recent years (Multimedia Appendix 6; left panel).

Figure 7. Trend charts of keywords. (A) Trend chart of author keywords. (B) Trend chart of Keywords Plus.



Through a deep analysis of the thematic map data derived from the author keywords (Figure 8A), this study gained insights into the potential evolutionary directions in the research field of breast cancer and ferroptosis. The upper right quadrant of the map features motor themes, which incorporate the pivotal thematic terms including ferroptosis, breast cancer, and its subtype, triple-negative breast cancer (TNBC), underscoring their interconnectedness and centrality in the research landscape. The niche themes in the upper left corner encompass thematic keyword clusters such as anticancer and gene signatures. The emerging or declining themes in the lower left corner include topics such as chemodynamic therapy and photothermal therapy (PTT). Finally, in the lower right corner, the basic themes include thematic keywords such as cancer microenvironment, immunotherapy, and prognosis model.

Figure 8. Thematic maps of keywords. (A) Thematic map of author keywords. (B) Thematic map of Keywords Plus.

Cluster Analysis of Keywords Plus and Analysis of the Main Evolution Trends in Research

The statistical data revealed that some papers had missing keywords, with an omission rate of 19.64%. By increasing the likelihood of yielding more precise outcomes, keywords derived using the Keywords Plus feature can enhance the rigor and reliability of research findings. Therefore, the following section presents a detailed analysis of Keywords Plus.

During the clustering analysis of Keywords Plus, this study excluded the core terms “breast cancer” and “ferroptosis,” along with their synonyms, focusing instead on the analysis of the top 100 most frequent terms. The analysis revealed that these terms could be grouped into five clusters (Multimedia Appendix 7). Among these clusters, the keyword “cell death” stands out as a focal point, with a high frequency of 99 occurrences. Additionally, “expression” (n=53), “mechanism” (n=49),

"transporter" (n=19), and "stress" (n=9) were identified as core terms within their respective clusters.

This paper further analyses the evolution trends of keywords in the past and future. As shown in Multimedia Appendix 6 (right panel) and Figure 7B, the results indicate that researchers in the field of breast cancer and ferroptosis have consistently maintained a certain level of interest in the study of cell death in recent years. Additionally, newly emerged keywords such as "mechanism," "expression," and "nanoparticle" have been frequently observed in recent years.

Through the visualization presented in the thematic map (Figure 8B), the results indicate that the motor themes in the upper right quadrant consist of keywords such as "migration," "epithelial-mesenchymal transition (EMT)," and "negative breast cancer." The niche themes in the upper left quadrant encompass two core themes: "radiotherapy" and "progression." Within the radiotherapy category, keywords such as "alpha" emerge, while in the progression category, keywords such as "discovery" and "hallmarks" appear. The emerging or declining themes in the lower left quadrant reveal keywords such as "transporter," "flavonoids," and "modes." The basic themes in the lower right quadrant exhibit two main themes: "expression" and "cell death." Within the expression category, keywords such as "apoptosis" and "ferroptosis" are present, while in the cell death category, keywords such as "mechanism" and "nanoparticle" emerge.

Discussion

Principal Findings

Breast cancer, as a malignancy with a high mortality rate, still faces significant limitations in early diagnosis and late-stage treatment. In recent years, ferroptosis, a unique form of cell death triggered by cellular metabolic imbalance, has attracted widespread attention [11], underscoring the importance and potential of research in this area. This study systematically reviewed 387 papers related to breast cancer and ferroptosis, retrieved from the WOSCC database, published between 2016 and February 1, 2024. The analysis revealed that the number of publications in this field peaked in 2023, showing a significant growth trend. However, the annual citation frequency of related publications exhibited fluctuations, with the highest citation rate observed in 2017, followed by a gradual decline. Further analysis highlighted the prominent contribution of scholars such as Efferth T, with Chen Y having the highest citation frequency. In terms of institutions, Sun Yat-sen University led in the number of publications. Research output from China and the United States dominated this field, with increasing collaborative efforts between the two countries. While China had the highest total citation count, Germany recorded the highest average citation frequency per paper. The journal *Frontiers in Oncology* published the highest number of papers. Citation network analysis indicated that journals in "Molecular, Biology, Genetics" were central to the citation network. High-impact papers focused on the induction mechanisms of ferroptosis in breast cancer and the regulatory role of related proteins. Keyword analysis identified GPX4 as a central research trend. Additionally, breast cancer subtypes emerged as key drivers of ferroptosis research. Fundamental research themes pointed to the tumor immune microenvironment, prognostic models, and nanotherapy as future research priorities.

In-Depth Analysis of Publication Volume and Citation Counts in Breast Cancer and Ferroptosis Research

A comprehensive analysis of breast cancer and ferroptosis research reveals that publication volume and citation counts serve as crucial indicators for assessing the impact and development of this field. These metrics not only reflect the significance of research outcomes but also highlight emerging trends. An examination of frequently cited papers shows that key studies focus on the induction mechanisms of ferroptosis in breast cancer cells, protein regulation, and comparisons with other forms of cell death, aiming to discover novel therapeutic strategies. Specifically, the first study in this field published in 2016 demonstrated the effectiveness of siramesine and lapatinib in inducing ferroptosis in breast cancer cells [12], leading to high citation rates. Since then, exploration of the relationship between breast cancer and ferroptosis has deepened, with research becoming increasingly refined. TNBC, notable research by Yu et al [13] revealed that exosome-encapsulated erastin elevates ferritin levels, suggesting a new therapeutic approach. Hasegawa et al [14] further elucidated the intricate interplay between the cystine/glutamate antiporter, CD44v, and the MUC1-C oncoprotein in TNBC, enhancing our understanding of this subtype. Additionally, studies on the glutathione (GSH)-degrading enzyme CHAC1 have uncovered a novel mechanism of ferroptosis in TNBC via the GCN2-eIF2α axis [15].

The analysis of influential authors highlights those with high publication volumes and citation impacts. A distinguished team led by Efferth T has investigated the cytotoxicity of natural products, such as alkaloids and saponins, against multidrug-resistant cancers, focusing on their ability to induce various forms of cell death, including ferroptosis [16-21]. Their work provides important insights into overcoming drug resistance and developing novel anticancer therapies. High-impact authors, such as Chen Y and Gibson SB, have explored the therapeutic potential of ferroptosis in breast cancer, particularly through nanotechnology and drug design, contributing new perspectives to the field [22,23].

Institutional analysis reveals that Sun Yat-sen University and Zhejiang University lead in publication output in this area, reflecting their strong research foundations. Journals such as *Frontiers in Oncology*, *ACS Nano*, and *Cell Death & Disease* play pivotal roles in disseminating key findings and shaping the breast cancer and ferroptosis research landscape. By understanding the complex relationships between citing and cited journals, we uncover the distribution characteristics of core journals and knowledge flow paths within this interdisciplinary field, highlighting the contributions of journals from molecular biology, immunology, genetics, physics, materials science, and chemistry.

At the national level, China and the United States dominate in publications related to breast cancer and ferroptosis. China has made significant contributions to exploring the application of ferroptosis in cancer treatment, particularly through drugs, gene

regulation, and nanotechnology. The United States, on the other hand, focuses on the mechanisms and regulation of ferroptosis in cancer, including its impact on lipid composition, iron metabolism, and key genes influencing ferroptosis sensitivity [23-27]. Despite limited resources, research teams from resource-poor countries like India and Brazil have leveraged nanotechnology to propose novel therapeutic strategies [28,29]. While China ranks first in total citations, Germany and the United States have higher citation rates per paper, indicating greater research impact and quality, possibly due to higher research investment and international collaboration.

The publication trend shows robust development in this field, with 28 publications by February 1, 2024, suggesting that 2024 may surpass 2023 in research output. The slight decline in citation rates reflects the surge in publications and underscores the growing interest in exploring the interplay between breast cancer and ferroptosis.

This growing research body highlights ferroptosis in breast cancer as an emerging frontier in biomedical science. Leading scholars in the field continue to make significant contributions, and institutions and countries demonstrate high levels of cooperation and investment. These positive factors predict that the field of breast cancer and ferroptosis research will achieve major breakthroughs in the near future, opening up new avenues for the diagnosis and treatment of breast cancer with profound implications.

A Systematic Summary of Ferroptosis Mechanisms

Ferroptosis, characterized by its unique mechanism involving iron-dependent regulation through multiple pathways, plays a pivotal role in the onset and progression of various diseases. The core mechanism of ferroptosis lies in the imbalance between phospholipid peroxidation triggered by the oxidation of polyunsaturated fatty acids and the antioxidant system mediated by GPX4. When GPX4 activity decreases or GSH levels, as a cofactor, decline, phospholipid peroxidation products accumulate, leading to ferroptosis [5]. Methods to induce ferroptosis include the use of ferroptosis inducers (such as erastin and RSL3), inhibition of the system Xc-/GSH-GPX4 axis, and increased iron accumulation. Conversely, strategies to inhibit ferroptosis involve upregulating GPX4 expression, using ferroptosis inhibitors (such as ferrostatin-1), and enhancing the antioxidant system [7].

The signaling pathways and defense mechanisms of ferroptosis are complex and diverse. The initiation of iron toxicity involves multiple sources of reactive oxygen species (ROS), including iron-mediated Fenton reactions, mitochondrial ROS, and membrane-associated ROS driven by the NADPH oxidase (NOX) protein family. Cells defend against ferroptosis by upregulating antioxidant gene expression, repairing membrane damage, and clearing damaged organelles [30]. There is also a crosstalk between the mitochondrial dynamic regulatory network and ferroptosis, with mitochondrial dysfunction and damage potentially promoting ferroptosis [31]. By regulating lipid metabolism pathways, ferroptosis can affect cellular sensitivity to ferroptosis, thereby modulating its occurrence and development [32]. Furthermore, mechanisms such as epigenetic modifications and posttranslational modifications also play crucial roles in the regulation of ferroptosis [33].

Given that iron accumulation, oxidative stress, and lipid peroxidation are common features of various diseases, ferroptosis, as a unique cell death mechanism, may be a key link in the pathological processes of diverse diseases [34]. In neurodegenerative diseases, such as Parkinson disease and ischemic stroke, ferroptosis also plays a significant role [35]. In the field of cancer, ferroptosis is considered a potential antitumor strategy. For example, ferroptosis interacts with the tumor immune microenvironment, influencing tumor growth and metastasis [36]. Inhibiting ubiquitin-specific protease 8 can destabilize GPX4, thereby enhancing cancer cell sensitivity to ferroptosis, inhibiting tumor growth, and enhancing the efficacy of immunotherapy [37]. Additionally, ferroptosis is closely related to the death of plaque cells (such as vascular endothelial cells, macrophages, and vascular smooth muscle cells) and the development of atherosclerotic plaques in cardiovascular diseases [38]. Therefore, the application of ferroptosis in the treatment of various diseases, including cancer immunotherapy and neurodegenerative diseases, is gaining increasing attention.

Research on the regulatory mechanisms of ferroptosis in breast cancer indicates that its regulation involves multiple aspects, such as lipid metabolism and the immune microenvironment [39-42], and that ferroptosis plays a crucial role in the sensitivity and drug resistance of breast cancer to therapy [43-45]. With a deeper understanding of ferroptosis mechanisms, the methods for its detection are continuously evolving, including fluorescence probes, positron emission tomography imaging, and others, providing multiscale perspectives for ferroptosis research [46].

Future research should further elucidate the specific mechanisms of ferroptosis in the onset and progression of diseases, as well as its specific roles in the regulatory pathways, sensitivity, and resistance of breast cancer, providing new targets and strategies for breast cancer treatment.

Keyword Analysis: Key Hotspot GPX4 and Ferroptosis Regulation in Breast Cancer

An analysis of evolutionary trends has revealed the increasingly prominent importance of GPX4 in relevant academic research fields. Previous studies have illustrated that GPX4, as a core regulatory factor of ferroptosis, plays a crucial role in organisms by catalyzing the reduction of phospholipid peroxides to inhibit iron ion degeneration, thereby affecting the ferroptosis process [47,48]. It has been reported that RUNX1 intron transcript 1 (RUNX1-it1) is significantly overexpressed in breast cancer tissues, and it can block the impact of ferroptosis by increasing GPX4 expression, thereby promoting the occurrence of breast cancer [49].

On the other hand, Tubastatin A, as an efficient and selective Histone Deacetylase 6 inhibitor, has been confirmed by relevant studies as a novel ferroptosis inducer. It promotes ferroptosis by inhibiting the activity of GPX4, thereby enhancing the effect of cancer radiotherapy [50]. Chloramine T has also been found to inhibit the cell growth of breast cancer and induce ferroptosis

by targeting the GPX4 axis, indicating that it may become a potential new strategy for breast cancer treatment [51].

In addition, recent research has provided new insights into the regulatory mechanism of GPX4. As a crucial selenoprotein, the function of GPX4 in resisting ferroptosis depends greatly on the presence of its unique selenocysteine residue, which provides an important basis for understanding the mechanism of GPX4 [52]. The process of selenocysteine uptake mediated by low-density lipoprotein receptor-related protein 8 has been elucidated as a core mechanism in promoting GPX4 protein synthesis and effectively inhibiting ferroptosis [53,54].

Another study found that phospholipid-modifying enzymes MBOAT1 and MBOAT2, as ferroptosis inhibitors, regulate ferroptosis by remodeling the cellular phospholipid profile through the sex hormone signaling pathway [55]. These findings indicate that research on GPX4 and its related regulatory mechanisms is gradually deepening from phenotypic description to molecular mechanism analysis, offering unprecedented opportunities for the treatment of breast cancer.

Motor Themes Analysis in Author Keywords: Differential Sensitivity of Ferroptosis in Various Breast Cancer Subtypes Driving the Field's Development

Breast cancer is a highly heterogeneous disease, with different breast cancer cell types, characterized by the expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and the proliferation marker Ki67, all exhibiting marked differences in sensitivity to ferroptosis. In the analysis of motor themes, findings related to ferroptosis sensitivity across breast cancer subtypes, including TNBC, have become well-established, closely interconnected with other emerging themes, and serve as a major driving force in the development of this field.

In ER-positive breast cancer, studies have shown that a combination of ferroptosis inducers, such as erastin and etoposide, can significantly enhance cellular sensitivity to ferroptosis by regulating IREB2/FPN1 expression and thus affecting iron metabolism, ultimately inducing ferroptosis [56]. Furthermore, ACSL4, a key regulator of ferroptosis, is upregulated in ER-positive breast cancer, promoting the accumulation of lipid peroxidation products and thereby enhancing cellular sensitivity to ferroptosis [57].

In contrast, sensitivity to ferroptosis in HER2-positive breast cancer may be regulated by multiple factors. For instance, inhibiting crVDAC3 can induce ferroptosis in breast cancer cells by reducing HSPB1 expression, thereby mediating resistance to deruxtecan in HER2-low breast cancer [58]. Additionally, m6A methylation modification of FGFR4 can inhibit ferroptosis, enhancing the resistance of HER2-positive breast cancer cells to anti-HER2 therapy [10]. Additionally, research has found that the inhibition of FGFR4 affects the expression of β-catenin/TCF4-SLC7A11 and FPN1, which has a significant impact on ferroptosis in HER2-positive breast cancer cells, thereby enhancing treatment sensitivity [10]. These studies suggest that the regulatory mechanisms of ferroptosis in HER2-positive breast cancer are more complex, involving the interaction of multiple signaling pathways.

In PR-positive breast cancer, although direct research on ferroptosis sensitivity is scarce, some indirect evidence suggests that PR may regulate the sensitivity of breast cancer cells to ferroptosis. Specifically, studies have reported that PR may indirectly regulate the sensitivity of TNBC to ferroptosis through metabolic pathways mediated by PR membrane component 1; however, these hypotheses await further experimental validation to establish their accuracy [59].

Relatively speaking, breast cancers that are negative for ER, PR, and HER2, especially TNBC, exhibit higher sensitivity to ferroptosis inducers. This characteristic may be closely related to the abnormal disruptions in iron and lipid metabolism in TNBC, which make TNBC cells more susceptible to ferroptosis inducers [60,61]. For example, the high expression level of ACSL4 in TNBC promotes the generation of lipid peroxides, a core step in the execution of ferroptosis [62]. Studies have also found that ferroptosis is enhanced through the combined action of GSH depletion and dihydroorotate dehydrogenase inhibitors [61]. Furthermore, the functions of key regulatory factors, such as the Xc-/GSH/GPX4 axis, ACSL4/LPCAT3 pathway, and nuclear factor erythroid 2-related factor 2 (NRF2) in ferroptosis, as well as their potential dysregulation mechanisms related to cancer cell survival and drug resistance, have been thoroughly explored [60]. Additionally, some natural compounds, such as artemisinin derivatives and traditional Chinese medicine extracts, have been found to inhibit the growth of TNBC cells by inducing ferroptosis [63].

Ki67, a commonly used proliferation marker, is often associated with higher proliferative activity and malignancy in breast cancer. Interestingly, breast cancer cells with high Ki67 expression also exhibit higher sensitivity to ferroptosis inducers [64]. However, research on the direct relationship between Ki67 expression and ferroptosis sensitivity is still limited. Future studies should further explore the role of Ki67 in the regulation of ferroptosis and whether it can serve as a biomarker for predicting the therapeutic effect of ferroptosis induction.

The motor themes analysis has provided a solid foundation for research in the area of breast cancer and ferroptosis. Continued investigation into the interactions between differential ferroptosis sensitivity across breast cancer subtypes holds the potential to uncover new therapeutic targets and biomarkers, ultimately driving the development of precision medicine and improving treatment outcomes and prognoses for patients with different breast cancer subtypes.

Analysis of Basic Themes in Breast Cancer and Ferroptosis Research: Immune Microenvironment and Prognostic Models Reveals Future Directions

The basic themes of immune microenvironment and prognostic models in breast cancer and ferroptosis research are closely interrelated and have led to several significant findings. For instance, in an insightful review by Shen et al [65], the authors explicitly point out the central role of iron in innate and adaptive immune responses, further emphasizing the potential positive influence of targeted regulation of iron metabolism on antitumor immunity and cancer treatment. Recent studies have demonstrated that T-cell–mediated immune responses [65] and

short-term acidosis-induced M1 macrophage polarization can effectively promote ferroptosis [66], thereby playing an active role in antitumor immunotherapy.

Research undertaken by Xu et al [67] has elegantly demonstrated that the development of ferroptosis-related gene (FRG) signatures holds significant promise as robust prognostic indicators for the immune microenvironment and therapeutic responsiveness. This approach shows great potential to inform and guide the design of future individualized treatment strategies, thereby enhancing precision medicine in breast cancer management.

A new lncRNA signal closely related to ferroptosis has been discovered, with its immune-related pathways significantly enriched in high-risk patients with breast cancer, demonstrating great potential for predicting the prognosis of patients with breast cancer [68]. Knocking down the hub gene SLC39A7, combined with TME scoring, can significantly affect the apoptosis and ferroptosis of cancer cells, with higher prognostic efficacy.

Meanwhile, epithelial cells and B cells exhibit higher ferroptosis scores, which are respectively related to the immune checkpoint blockade response and immune checkpoint blockade-independent gene expression, suggesting a connection between ferroptosis and the immune microenvironment [69]. It is noteworthy that research has accurately predicted the survival of patients with breast cancer through a novel prognostic model composed of nine FRGs [70].

Chen et al [71] used the Cancer Genome Atlas to identify 11 ferroptosis genes related to breast cancer prognosis, established a precise prognostic model, and analyzed therapeutic targets, thereby enhancing prognosis management. In TNBC, the subtype of breast cancer with the worst prognosis, research by Wu et al [72] evaluated the role of FRGs and found that they may affect prognostic models by regulating the tumor microenvironment, thus providing a solid theoretical basis for a clinically accurate prediction of TNBC prognosis. Furthermore, studies have evaluated the FRG AKR1C1, revealing its potential related to the immune-microenvironment, which may further affect the progress and prognosis of patients with breast cancer, making it a novel biomarker for the immune microenvironment and prognosis determination of breast cancer [73].

In conclusion, although significant progress has been made in the areas of immune microenvironment and prognostic models within the context of breast cancer and ferroptosis research, these fields remain underexplored and offer considerable potential for future investigation. These themes provide valuable directions for further research and offer essential guidance for researchers aiming to deepen their work in this area.

Innovative Applications and Strategies of Nanoparticles: Further Supplementing Future Research Directions in Breast Cancer and Ferroptosis

Through an in-depth analysis of the Keywords Plus basic themes, we further explore the future development directions of nanoparticles in the context of breast cancer and ferroptosis research. The application of nanoparticles in breast cancer treatment relies heavily on their unique physical and chemical properties to achieve the targeted delivery and controlled release of drugs, thereby enhancing therapeutic efficacy and reducing side effects [74]. Nanoparticles can exert their effects through various mechanisms that promote cell death, including ferroptosis [75]. Recent research has reported the use of a novel Fenton-independent pathway using photothermal nanozymes to overcome the limitations of the Fenton reaction. By using iron-containing hollow mesoporous Prussian blue (HMPB) nanocubes as an iron source to prepare HMPB@Lip, which combines PTT with nanozyme action, the induced ferroptosis through a non-Fenton reaction effectively ablates tumors, demonstrating the therapeutic potential of HMPB@Lip as a multifunctional nanozyme for ferroptosis. Furthermore, the development of Her2-DSG NPs nanostructures, which combine chemo-PTT with tumor microenvironment remodeling and immune activation, holds great promise for HER2-positive breast cancer [76].

Another study constructed nFeAPG nano-complex synergizing Fe^{3} with apigenin to enhance immune response through PTT, effectively controlling TNBC [71]. Additionally, the ICG@SANPs-cRGD nanomaterial-mediated combination of PTT/photodynamic therapy and immunotherapy has demonstrated its potential as a multifunctional platform for breast cancer treatment [77].

Furthermore, a search conducted on the Clinical Trials website revealed that there is only one clinical trial project related to breast cancer and ferroptosis, and that it is associated with nanocarriers (Multimedia Appendix 8). As of the search date, November 12, 2024, this project is still in the recruiting phase. Its focus is on the application of carbon nanoparticle-loaded iron (CNSI-Fe (II)) in the treatment of patients with advanced solid tumors, particularly those who have experienced treatment failure (including disease progression or intolerance) or lack of available standard therapeutic options (trial identifier: NCT06048367). The primary end point of this study is to assess the safety and tolerability of CNSI-Fe (II) throughout the study duration. The secondary end point, meanwhile, focuses on evaluating the pharmacokinetic profile of CNSI-Fe (II) in patients with advanced solid tumors.

Nanotechnology, as a fundamental topic in the research fields of breast cancer and ferroptosis, is still in its relatively early stages of development. Nonetheless, these research achievements have not only significantly expanded the application scope of nanotechnology in breast cancer treatment but have also highlighted its close connection with breast cancer therapy and identified numerous promising therapeutic research directions that await further exploration. More importantly, they have pointed out new avenues for future research endeavors and provided invaluable guidance for investigative pursuits.

Limitations

This study represents the first comprehensive bibliometric analysis of 387 academic papers related to breast cancer and ferroptosis sourced from the WOSCC database. By applying multiple analytical approaches, it thoroughly investigates the research hotspots and future trends in this field, offering a

comprehensive and systematic overview of the latest developments in this area. These strengths significantly enhance the academic value and practical relevance of the study, providing essential guidance for future advancements in the field of breast cancer and ferroptosis.

However, several limitations of our methodology must be acknowledged. First, although the WOSCC database is vast, it was not specifically designed for bibliometric analysis and relies on a single data source, which may not capture all relevant literature present in other databases (such as PubMed, Scopus, and Google Scholar), potentially introducing bias. Future research should integrate multiple databases to mitigate this issue. Second, our focus on English-language publications limits the comprehensiveness of the study by excluding high-quality non-English research. To address this, future studies should adopt multilingual search strategies. Third, the exclusion of nonpaper publications (eg, reviews and conference papers) constrains the scope of research outputs considered. To obtain a more holistic view, future work should adopt more inclusive selection criteria. Finally, given the rapid updates in databases, there is a risk of overlooking recent advances in the field's literature. Periodic updates to bibliometric analyses are recommended to ensure timely and accurate capture of the latest trends and developments in breast cancer and ferroptosis research.

Conclusions

Ferroptosis holds significant value and potential in breast cancer research. Through bibliometric analysis, this study suggests that enhancing collaboration and communication among authors, institutions, and countries, while leveraging the interdisciplinary nature of journal citations, can facilitate advances in fundamental research on ferroptosis in breast cancer. Specifically, this paper reviews the progress in understanding the mechanisms of ferroptosis, providing guidance for breast cancer research. It emphasizes the sensitivity of different breast cancer subtypes to ferroptosis as a driving force in the current research landscape and highlights the pivotal role of GPX4 in the regulatory mechanisms of ferroptosis in breast cancer as a key focus. Additionally, the study identifies the importance of the tumor microenvironment, immunotherapy, and prognostic models for future development. Although the application of nanoparticles in breast cancer treatment is in its preliminary stages, it has already shown promising prospects and serves as a rich supplement to the future directions. By elucidating the current status and future trends of ferroptosis in breast cancer research, this study provides direction for future research endeavors and promotes innovation in breast cancer treatment strategies.

Acknowledgments

All authors would like to thank the "Guangxi Zhuang Autonomous Region Clinical Medicine Research Center for Molecular Pathology and Intelligent Pathology Precision Diagnosis" for providing relevant technical support. This study was supported by the Guangxi Medical University Teacher Teaching Ability Development Project (2202JFA02), and the Guangxi Medical University Undergraduate Education and Teaching Reform Project (2023Y05).

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JYL, YLD, SYL, SYC, RQH, DYQ, BTC, GC, XY, and WP contributed to the study conception and design. Material preparation, data collection, and analysis were performed by JYL, YLD, SYL, SYC, DYQ, and BTC. The first draft of the manuscript was written by JYL and YLD. SYL, RQH, GC, XY, and WP revised the paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key information of the included studies.
[PNG File , 377 KB - ijmr_v14i1e66286_app1.png]

Multimedia Appendix 2
Top 10 most frequently cited articles in research on ferroptosis and breast cancer.
[XLSX File (Microsoft Excel File), 11 KB - ijmr_v14i1e66286_app2.xlsx]

Multimedia Appendix 3
Top 10 authors with the most citations in ferroptosis and breast cancer research within the field.
[XLSX File (Microsoft Excel File), 276 KB - ijmr_v14i1e66286_app3.xlsx]

Multimedia Appendix 4
Top 10 countries ranked by total citations and average article citations in the field of ferroptosis and breast cancer research.
[XLSX File (Microsoft Excel File), 10 KB - ijmr_v14i1e66286_app4.xlsx]

Multimedia Appendix 5
Author keywords clustering visualisation in the field of breast cancer and ferroptosis research.
[PNG File , 915 KB - ijmr_v14i1e66286_app5.png]

Multimedia Appendix 6
Heatmap of keyword analysis.
[PNG File , 742 KB - ijmr_v14i1e66286_app6.png]

Multimedia Appendix 7
Visualisation of keywords plus clustering in breast cancer and ferroptosis research fields.
[PNG File , 880 KB - ijmr_v14i1e66286_app7.png]

Multimedia Appendix 8
Summary of clinical trials in the field of ferroptosis and breast cancer research (n=1).
[XLSX File (Microsoft Excel File), 13 KB - ijmr_v14i1e66286_app8.xlsx]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast 2022;66:15-23 [FREE Full text] [doi: 10.1016/j.breast.2022.08.010] [Medline: 36084384]
3. Admoun C, Mayrovitz HN. The eiology of breast cancer. In: Mayrovitz HN, editor. Breast Cancer [Internet]. Brisbane (AU): Exon Publications; 2022.
4. Breast Cancer Expert Committee of National Cancer Quality Control Center, Breast Cancer Expert Committee of China Anti-Cancer Association, Cancer Drug Clinical Research Committee of China Anti-Cancer Association. Guidelines for clinical diagnosis and treatment of advanced breast cancer in China (2022 edition). Zhonghua Zhong Liu Za Zhi 2022;44(12):1262-1287. [doi: 10.3760/cma.j.cn112152-20221007-00680] [Medline: 36575780]
5. Jiang X, Stockwell BR, Conrad M. Ferroptosis: mechanisms, biology and role in disease. Nat Rev Mol Cell Biol 2021;22(4):266-282 [FREE Full text] [doi: 10.1038/s41580-020-00324-8] [Medline: 33495651]
6. Li J, Cao F, Yin HL, Huang ZJ, Lin ZT, Mao N, et al. Ferroptosis: past, present and future. Cell Death Dis 2020;11(2):88 [FREE Full text] [doi: 10.1038/s41419-020-2298-2] [Medline: 32015325]
7. Tang D, Chen X, Kang R, Kroemer G. Ferroptosis: molecular mechanisms and health implications. Cell Res 2021;31(2):107-125 [FREE Full text] [doi: 10.1038/s41422-020-00441-1] [Medline: 33268902]
8. Dong X, Li Y, Sheng X, Zhou W, Sun A, Dai H. Mitochondria-related signaling pathways involved in breast cancer regulate ferroptosis. Genes Dis 2024;11(1):358-366 [FREE Full text] [doi: 10.1016/j.gendis.2023.03.019] [Medline: 37588231]
9. Xu Z, Wang X, Sun W, Xu F, Kou H, Hu W, et al. RelB-activated GPX4 inhibits ferroptosis and confers tamoxifen resistance in breast cancer. Redox Biol 2023;68:102952 [FREE Full text] [doi: 10.1016/j.redox.2023.102952] [Medline: 37944384]
10. Zou Y, Zheng S, Xie X, Ye F, Hu X, Tian Z, et al. N6-methyladenosine regulated FGFR4 attenuates ferroptotic cell death in recalcitrant HER2-positive breast cancer. Nat Commun 2022;13(1):2672 [FREE Full text] [doi: 10.1038/s41467-022-30217-7] [Medline: 35562334]
11. Mao C, Wang M, Zhuang L, Gan B. Metabolic cell death in cancer: ferroptosis, cuproptosis, disulfidptosis, and beyond. Protein Cell 2024;15(9):642-660 [FREE Full text] [doi: 10.1093/procel/pwae003] [Medline: 38428031]
12. Ma S, Henson ES, Chen Y, Gibson SB. Ferroptosis is induced following siramesine and lapatinib treatment of breast cancer cells. Cell Death Dis 2016;7(7):e2307 [FREE Full text] [doi: 10.1038/cddis.2016.208] [Medline: 27441659]
13. Yu M, Gai C, Li Z, Ding D, Zheng J, Zhang W, et al. Targeted exosome-encapsulated erastin induced ferroptosis in triple negative breast cancer cells. Cancer Sci 2019;110(10):3173-3182 [FREE Full text] [doi: 10.1111/cas.14181] [Medline: 31464035]
14. Hasegawa M, Takahashi H, Rajabi H, Alam M, Suzuki Y, Yin L, et al. Functional interactions of the cystine/glutamate antiporter, CD44v and MUC1-C oncoprotein in triple-negative breast cancer cells. Oncotarget 2016;7(11):11756-11769 [FREE Full text] [doi: 10.18632/oncotarget.7598] [Medline: 26930718]

15. Chen MS, Wang SF, Hsu CY, Yin PH, Yeh TS, Lee HC, et al. CHAC1 degradation of glutathione enhances cystine-starvation-induced necroptosis and ferroptosis in human triple negative breast cancer cells via the GCN2-eIF2α-ATF4 pathway. Oncotarget 2017;8(70):114588-114602 [FREE Full text] [doi: 10.18632/oncotarget.23055] [Medline: 29383104]
16. Mbaveng AT, Noulala CGT, Samba ARM, Tankeo SB, Abdelfatah S, Fotso GW, et al. The alkaloid, soyauxinium chloride, displays remarkable cytotoxic effects towards a panel of cancer cells, inducing apoptosis, ferroptosis and necroptosis. Chem Biol Interact 2021;333:109334. [doi: 10.1016/j.cbi.2020.109334] [Medline: 33245930]
17. Mbaveng AT, Bitchagno GTM, Kuete V, Tane P, Efferth T. Cytotoxicity of ungeremine towards multi-factorial drug resistant cancer cells and induction of apoptosis, ferroptosis, necroptosis and autophagy. Phytomedicine 2019;60:152832. [doi: 10.1016/j.phymed.2019.152832] [Medline: 31031043]
18. Mbaveng AT, Fotso GW, Ngnintedo D, Kuete V, Ngadjui BT, Keumedjio F, et al. Cytotoxicity of epunctanone and four other phytochemicals isolated from the medicinal plants Garcinia epunctata and Ptycholobium contortum towards multi-factorial drug resistant cancer cells. Phytomedicine 2018;48:112-119. [doi: 10.1016/j.phymed.2017.12.016] [Medline: 30195869]
19. Mbaveng AT, Ndontsa BL, Kuete V, Nguekeu YMM, Çelik İ, Mbouangouere R, et al. A naturally occuring triterpene saponin ardisiacrispin B displayed cytotoxic effects in multi-factorial drug resistant cancer cells via ferroptotic and apoptotic cell death. Phytomedicine 2018;43:78-85. [doi: 10.1016/j.phymed.2018.03.035] [Medline: 29747757]
20. Mbaveng AT, Chi GF, Bonsou IN, Abdelfatah S, Tamfu AN, Yeboah EMO, et al. N-acetylglycoside of oleanolic acid (aridanin) displays promising cytotoxicity towards human and animal cancer cells, inducing apoptotic, ferroptotic and necroptotic cell death. Phytomedicine 2020;76:153261. [doi: 10.1016/j.phymed.2020.153261] [Medline: 32559584]
21. Mbaveng AT, Damen F, Guefack MF, Tankeo SB, Abdelfatah S, Bitchagno GTM, et al. 8,8-bis-(Dihydroconiferyl)-diferulate displayed impressive cytotoxicity towards a panel of human and animal cancer cells. Phytomedicine 2020;70:153215. [doi: 10.1016/j.phymed.2020.153215] [Medline: 32388040]
22. Ma S, Dielschneider RF, Henson ES, Xiao W, Choquette TR, Blankstein AR, et al. Ferroptosis and autophagy induced cell death occur independently after siramesine and lapatinib treatment in breast cancer cells. PLoS One 2017;12(8):e0182921 [FREE Full text] [doi: 10.1371/journal.pone.0182921] [Medline: 28827805]
23. Chen Y, Yao Z, Liu P, Hu Q, Huang Y, Ping L, et al. A self-assembly nano-prodrug for triple-negative breast cancer combined treatment by ferroptosis therapy and chemotherapy. Acta Biomater 2023;159:275-288. [doi: 10.1016/j.actbio.2023.01.050] [Medline: 36709836]
24. Yi J, Zhu J, Wu J, Thompson CB, Jiang X. Oncogenic activation of PI3K-AKT-mTOR signaling suppresses ferroptosis via SREBP-mediated lipogenesis. Proc Natl Acad Sci U S A 2020;117(49):31189-31197 [FREE Full text] [doi: 10.1073/pnas.2017152117] [Medline: 33229547]
25. Brown CW, Amante JJ, Chhoy P, Elaimy AL, Liu H, Zhu LJ, et al. Prominin2 drives ferroptosis resistance by stimulating iron export. Dev Cell 2019;51(5):575-586.e4 [FREE Full text] [doi: 10.1016/j.devcel.2019.10.007] [Medline: 31735663]
26. Liu T, Jiang L, Tavana O, Gu W. The deubiquitylase OTUB1 mediates ferroptosis via stabilization of SLC7A11. Cancer Res 2019;79(8):1913-1924 [FREE Full text] [doi: 10.1158/0008-5472.CAN-18-3037] [Medline: 30709928]
27. Llabani E, Hicklin RW, Lee HY, Motika SE, Crawford LA, Weerapana E, et al. Diverse compounds from pleuromutilin lead to a thioredoxin inhibitor and inducer of ferroptosis. Nat Chem 2019;11(6):521-532 [FREE Full text] [doi: 10.1038/s41557-019-0261-6] [Medline: 31086302]
28. Szwed M, Torgersen ML, Kumari RV, Yadava SK, Pust S, Iversen TG, et al. Biological response and cytotoxicity induced by lipid nanocapsules. J Nanobiotechnol 2020;18(1):5 [FREE Full text] [doi: 10.1186/s12951-019-0567-y] [Medline: 31907052]
29. Zhu J, Dai P, Liu F, Li Y, Qin Y, Yang Q, et al. Upconverting nanocarriers enable triggered microtubule inhibition and concurrent ferroptosis induction for selective treatment of triple-negative breast cancer. Nano Lett 2020;20(9):6235-6245. [doi: 10.1021/acs.nanolett.0c00502] [Medline: 32804509]
30. Liu J, Kang R, Tang D. Signaling pathways and defense mechanisms of ferroptosis. FEBS J 2022;289(22):7038-7050 [FREE Full text] [doi: 10.1111/febs.16059] [Medline: 34092035]
31. Li J, Jia YC, Ding YX, Bai J, Cao F, Li F. The crosstalk between ferroptosis and mitochondrial dynamic regulatory networks. Int J Biol Sci 2023;19(9):2756-2771 [FREE Full text] [doi: 10.7150/ijbs.83348] [Medline: 37324946]
32. Pope LE, Dixon SJ. Regulation of ferroptosis by lipid metabolism. Trends Cell Biol 2023;33(12):1077-1087. [doi: 10.1016/j.tcb.2023.05.003] [Medline: 37407304]
33. Wang Y, Hu J, Wu S, Fleishman JS, Li Y, Xu Y, et al. Targeting epigenetic and posttranslational modifications regulating ferroptosis for the treatment of diseases. Signal Transduction Targeted Ther 2023;8(1):449 [FREE Full text] [doi: 10.1038/s41392-023-01720-0] [Medline: 38072908]
34. Mahoney-Sánchez L, Bouchaoui H, Ayton S, Devos D, Duce JA, Devedjian JC. Ferroptosis and its potential role in the physiopathology of Parkinson's disease. Prog Neurobiol 2021;196:101890 [FREE Full text] [doi: 10.1016/j.pneurobio.2020.101890] [Medline: 32726602]
35. Hu X, Bao Y, Li M, Zhang W, Chen C. The role of ferroptosis and its mechanism in ischemic stroke. Exp Neurol 2024;372:114630. [doi: 10.1016/j.expneurol.2023.114630] [Medline: 38056585]

36. Chen X, Kang R, Kroemer G, Tang D. Broadening horizons: the role of ferroptosis in cancer. Nat Rev Clin Oncol 2021;18(5):280-296. [doi: 10.1038/s41571-020-00462-0] [Medline: 33514910]
37. Li H, Sun Y, Yao Y, Ke S, Zhang N, Xiong W, et al. USP8-governed GPX4 homeostasis orchestrates ferroptosis and cancer immunotherapy. Proc Natl Acad Sci U S A 2024;121(16):e2315541121. [doi: 10.1073/pnas.2315541121] [Medline: 38598341]
38. Xu X, Xu XD, Ma MQ, Liang Y, Cai YB, Zhu ZX, et al. The mechanisms of ferroptosis and its role in atherosclerosis. Biomed Pharmacother 2024;171:116112 [FREE Full text] [doi: 10.1016/j.biopha.2023.116112] [Medline: 38171246]
39. Ding Y, Chen X, Liu C, Ge W, Wang Q, Hao X, et al. Identification of a small molecule as inducer of ferroptosis and apoptosis through ubiquitination of GPX4 in triple negative breast cancer cells. J Hematol Oncol 2021;14(1):19 [FREE Full text] [doi: 10.1186/s13045-020-01016-8] [Medline: 33472669]
40. Wang Z, Li R, Hou N, Zhang J, Wang T, Fan P, et al. PRMT5 reduces immunotherapy efficacy in triple-negative breast cancer by methylating KEAP1 and inhibiting ferroptosis. J Immunother Cancer 2023;11(6):e006890 [FREE Full text] [doi: 10.1136/jitc-2023-006890] [Medline: 37380368]
41. Doll S, Proneth B, Tyurina YY, Panzilius E, Kobayashi S, Ingold I, et al. ACSL4 dictates ferroptosis sensitivity by shaping cellular lipid composition. Nat Chem Biol 2017;13(1):91-98 [FREE Full text] [doi: 10.1038/nchembio.2239] [Medline: 27842070]
42. Zhao Y, Liu Z, Liu G, Zhang Y, Liu S, Gan D, et al. Neutrophils resist ferroptosis and promote breast cancer metastasis through aconitate decarboxylase 1. Cell Metab 2023;35(10):1688-1703.e10 [FREE Full text] [doi: 10.1016/j.cmet.2023.09.004] [Medline: 37793345]
43. Shen M, Cao S, Long X, Xiao L, Yang L, Zhang P, et al. DNAJC12 causes breast cancer chemotherapy resistance by repressing doxorubicin-induced ferroptosis and apoptosis via activation of AKT. Redox Biol 2024;70:103035 [FREE Full text] [doi: 10.1016/j.redox.2024.103035] [Medline: 38306757]
44. Yang J, Zhou Y, Xie S, Wang J, Li Z, Chen L, et al. Metformin induces ferroptosis by inhibiting UFMylation of SLC7A11 in breast cancer. J Exp Clin Cancer Res 2021;40(1):206 [FREE Full text] [doi: 10.1186/s13046-021-02012-7] [Medline: 34162423]
45. Zhu Z, Shen H, Xu J, Fang Z, Wo G, Ma Y, et al. GATA3 mediates doxorubicin resistance by inhibiting CYB5R2-catalyzed iron reduction in breast cancer cells. Drug Resist Updates 2023;69:100974 [FREE Full text] [doi: 10.1016/j.drup.2023.100974] [Medline: 37230023]
46. Zeng F, Nijiati S, Tang L, Ye J, Zhou Z, Chen X. Ferroptosis detection: from approaches to applications. Angew Chem Int Ed Engl 2023;62(35):e202300379. [doi: 10.1002/anie.202300379] [Medline: 36828775]
47. Liang D, Feng Y, Zandkarimi F, Wang H, Zhang Z, Kim J, et al. Ferroptosis surveillance independent of GPX4 and differentially regulated by sex hormones. Cell 2023;186(13):2748-2764.e22 [FREE Full text] [doi: 10.1016/j.cell.2023.05.003] [Medline: 37267948]
48. Luo M, Yan J, Hu X, Li H, Li H, Liu Q, et al. Targeting lipid metabolism for ferroptotic cancer therapy. Apoptosis 2023;28(1-2):81-107. [doi: 10.1007/s10495-022-01795-0] [Medline: 36399287]
49. Wang S, Wang Y, Li Q, Zeng K, Li X, Feng X. RUNX1-IT1 favors breast cancer carcinogenesis through regulation of IGF2BP1/GPX4 axis. Discov Oncol 2023;14(1):42 [FREE Full text] [doi: 10.1007/s12672-023-00652-z] [Medline: 37036576]
50. Liu S, Zhang HL, Li J, Ye ZP, Du T, Li LC, et al. Tubastatin a potently inhibits GPX4 activity to potentiate cancer radiotherapy through boosting ferroptosis. Redox Biol 2023;62:102677 [FREE Full text] [doi: 10.1016/j.redox.2023.102677] [Medline: 36989572]
51. Li H, Liu W, Zhang X, Wu F, Sun D, Wang Z. Ketamine suppresses proliferation and induces ferroptosis and apoptosis of breast cancer cells by targeting KAT5/GPX4 axis. Biochem Biophys Res Commun 2021;585:111-116. [doi: 10.1016/j.bbrc.2021.11.029] [Medline: 34800882]
52. Ingold I, Berndt C, Schmitt S, Doll S, Poschmann G, Buday K, et al. Selenium utilization by GPX4 is required to prevent hydroperoxide-induced ferroptosis. Cell 2018;172(3):409-422.e21 [FREE Full text] [doi: 10.1016/j.cell.2017.11.048] [Medline: 29290465]
53. Alborzinia H, Chen Z, Yildiz U, Freitas FP, Vogel FCE, Varga JP, et al. LRP8-mediated selenocysteine uptake is a targetable vulnerability in MYCN-amplified neuroblastoma. EMBO Mol Med 2023;15(8):e18014 [FREE Full text] [doi: 10.15252/emmm.202318014] [Medline: 37435859]
54. Li Z, Ferguson L, Deol KK, Roberts MA, Magtanong L, Hendricks JM, et al. Ribosome stalling during selenoprotein translation exposes a ferroptosis vulnerability. Nat Chem Biol 2022;18(7):751-761 [FREE Full text] [doi: 10.1038/s41589-022-01033-3] [Medline: 35637349]
55. NA. Sex hormone signaling suppresses ferroptosis via phospholipid remodeling. Cancer Discovery 2023;13(8):1759. [doi: 10.1158/2159-8290.CD-RW2023-091] [Medline: 37326388]
56. Ozkan E, Bakar-Ates F. Etoposide in combination with erastin synergistically altered iron homeostasis and induced ferroptotic cell death through regulating IREB2/FPN1 expression in estrogen receptor positive-breast cancer cells. Life Sci 2023;312:121222. [doi: 10.1016/j.lfs.2022.121222] [Medline: 36442526]

57. Hou J, Jiang C, Wen X, Li C, Xiong S, Yue T, et al. ACSL4 as a potential target and biomarker for anticancer: from molecular mechanisms to clinical therapeutics. Front Pharmacol 2022;13:949863 [FREE Full text] [doi: 10.3389/fphar.2022.949863] [Medline: 35910359]
58. Zou Y, Yang A, Chen B, Deng X, Xie J, Dai D, et al. crVDAC3 alleviates ferroptosis by impeding HSPB1 ubiquitination and confers trastuzumab deruxtecan resistance in HER2-low breast cancer. Drug Resist Updates 2024;77:101126. [doi: 10.1016/j.drup.2024.101126] [Medline: 39243601]
59. Zhao Y, Ruan X, Cheng J, Xu X, Gu M, Mueck AO. PGRMC1 promotes triple-negative breast cancer cell growth via suppressing ferroptosis. Climacteric 2023;26(2):135-142. [doi: 10.1080/13697137.2023.2170225] [Medline: 36724820]
60. Mokhtarpour K, Razi S, Rezaei N. Ferroptosis as a promising targeted therapy for triple negative breast cancer. Breast Cancer Res Treat 2024;207(3):497-513. [doi: 10.1007/s10549-024-07387-7] [Medline: 38874688]
61. Zhou TJ, Zhang MM, Liu DM, Huang LL, Yu HQ, Wang Y, et al. Glutathione depletion and dihydroorotate dehydrogenase inhibition actuated ferroptosis-augment to surmount triple-negative breast cancer. Biomaterials 2024;305:122447. [doi: 10.1016/j.biomaterials.2023.122447] [Medline: 38154441]
62. Monaco ME. ACSL4: biomarker, mediator and target in quadruple negative breast cancer. Oncotarget 2023;14:563-575 [FREE Full text] [doi: 10.18632/oncotarget.28453] [Medline: 37306503]
63. Wang Y, Yuan X, Ren M, Wang Z. Ferroptosis: a new research direction of artemisinin and its derivatives in anti-cancer treatment. Am J Chin Med 2024;52(1):161-181. [doi: 10.1142/S0192415X24500071] [Medline: 38328829]
64. Yeh NT, Lin TC, Liu IJ, Hu SH, Hsu TC, Chin HY, et al. Hyaluronic acid-covered ferric ion-rich nanobullets with high zoledronic acid payload for breast tumor-targeted chemo/chemodynamic therapy. Int J Biol Macromol 2024;279(Pt 3):135271. [doi: 10.1016/j.ijbiomac.2024.135271] [Medline: 39233170]
65. Shen L, Zhou Y, He H, Chen W, Lenahan C, Li X, et al. Crosstalk between macrophages, T cells, and iron metabolism in tumor microenvironment. Oxid Med Cell Longevity 2021;2021:8865791 [FREE Full text] [doi: 10.1155/2021/8865791] [Medline: 33628389]
66. Xiong H, Zhai Y, Meng Y, Wu Z, Qiu A, Cai Y, et al. Acidosis activates breast cancer ferroptosis through ZFAND5/SLC3A2 signaling axis and elicits M1 macrophage polarization. Cancer Lett 2024;587:216732. [doi: 10.1016/j.canlet.2024.216732] [Medline: 38360142]
67. Xu Y, Du Y, Zheng Q, Zhou T, Ye B, Wu Y, et al. Identification of ferroptosis-related prognostic signature and subtypes related to the immune microenvironment for breast cancer patients receiving neoadjuvant chemotherapy. Front Immunol 2022;13:895110 [FREE Full text] [doi: 10.3389/fimmu.2022.895110] [Medline: 35603151]
68. Zhang K, Ping L, Du T, Liang G, Huang Y, Li Z, et al. A ferroptosis-related lncRNAs signature predicts prognosis and immune microenvironment for breast cancer. Front Mol Biosci 2021;8:678877 [FREE Full text] [doi: 10.3389/fmolb.2021.678877] [Medline: 34164433]
69. Qiu Z, Deng C, Zhou F, Chen Y, Chen X, Liu X, et al. Ferroptosis heterogeneity within the tumor microenvironment revealed a genetic blueprint of breast cancer. Environ Toxicol 2024;39(5):2741-2752. [doi: 10.1002/tox.24142] [Medline: 38251953]
70. Lu YJ, Gong YJ, Li WY, Zhao C, Guo F. The prognostic significance of a novel ferroptosis-related gene model in breast cancer. Ann Transl Med 2022;10(4):184 [FREE Full text] [doi: 10.21037/atm-22-479] [Medline: 35280394]
71. Chen R, Jiang Z, Cheng Y, Ye J, Li S, Xu Y, et al. Multifunctional iron-apigenin nanocomplex conducting photothermal therapy and triggering augmented immune response for triple negative breast cancer. Int J Pharm 2024;655:124016. [doi: 10.1016/j.ijpharm.2024.124016] [Medline: 38503397]
72. Wu S, Pan R, Lu J, Wu X, Xie J, Tang H, et al. Development and verification of a prognostic ferroptosis-related gene model in triple-negative breast cancer. Front Oncol 2022;12:896927 [FREE Full text] [doi: 10.3389/fonc.2022.896927] [Medline: 35719954]
73. Zhang Z, Qiu X, Yan Y, Liang Q, Cai Y, Peng B, et al. Evaluation of ferroptosis-related gene as a novel biomarker associated with the immune microenvironment and prognosis in breast cancer. Int J Gen Med 2021;14:6189-6200 [FREE Full text] [doi: 10.2147/IJGM.S329031] [Medline: 34611432]
74. Marshall SK, Angsantikul P, Pang Z, Nasongkla N, Hussen RSD, Thamphiwatana SD. Biomimetic targeted theranostic nanoparticles for breast cancer treatment. Molecules 2022;27(19):6473 [FREE Full text] [doi: 10.3390/molecules27196473] [Medline: 36235009]
75. Herdiana Y, Sriwidodo S, Sofian FF, Wilar G, Diantini A. Nanoparticle-based antioxidants in stress signaling and programmed cell death in breast cancer treatment. Molecules 2023;28(14):5305 [FREE Full text] [doi: 10.3390/molecules28145305] [Medline: 37513179]
76. Zheng D, Wan C, Yang H, Xu L, Dong Q, Du C, et al. Her2-targeted multifunctional nano-theranostic platform mediates tumor microenvironment remodeling and immune activation for breast cancer treatment. Int J Nanomed 2020;15:10007-10028 [FREE Full text] [doi: 10.2147/IJN.S271213] [Medline: 33376321]
77. Kong C, Xu B, Qiu G, Wei M, Zhang M, Bao S, et al. Multifunctional nanoparticles-mediated PTT/PDT synergistic immune activation and antitumor activity combined with anti-PD-L1 immunotherapy for breast cancer treatment. Int J Nanomed 2022;17:5391-5411 [FREE Full text] [doi: 10.2147/IJN.S373282] [Medline: 36419717]

Abbreviations

ER: estrogen receptor
FRG: ferroptosis-related gene
GPX4: glutathione peroxidase 4
GSH: glutathione
HER2: human epidermal growth factor receptor 2
HMPB: hollow mesoporous Prussian blue
PPT: photothermal therapy
PR: progesterone receptor
ROS: reactive oxygen species
TNBC: triple-negative breast cancer
WOSCC: Web of Science Core Collection

Edited by T de Azevedo Cardoso; submitted 09.09.24; peer-reviewed by Y Zou, C Mao, Z Li; comments to author 01.10.24; revised version received 26.11.24; accepted 05.12.24; published 26.02.25.

Please cite as:
Luo JY, Deng YL, Lu SY, Chen SY, He RQ, Qin DY, Chi BT, Chen G, Yang X, Peng W
Current Status and Future Directions of Ferroptosis Research in Breast Cancer: Bibliometric Analysis
Interact J Med Res 2025;14:e66286
URL: https://www.i-jmr.org/2025/1/e66286
doi:10.2196/66286
PMID:

Review

Process Evaluations of Interventions for the Prevention of Type 2 Diabetes in Women With Gestational Diabetes Mellitus: Systematic Review

Iklil Iman Mohd Sa'id[1*], PhD; Natasha Hotung[2*], BA(Hons), MA, MSc; Madeleine Benton[2], CPychol, PhD; Iliatha Papachristou Nadal[2], CPsychol, PhD; Anisah Baharom[1], PhD; Matthew Prina[3], PhD; Barakatun Nisak Mohd Yusof[1], PhD; Kimberley Goldsmith[2], PhD; Samantha Birts[4], BSc; Ching Siew Mooi[1], MD, MMed Fam, PhD; Angus Forbes[2], BSc, MSc, PhD; Khalida Ismail[2], MBBS, PhD; Boon How Chew[1], MD, PhD

[1]Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia

[2]Department of Psychological Medicine, King's College London, London, United Kingdom

[3]Newcastle University, Newcastle upon Tyne, United Kingdom

[4]Department of Psychological Medicine, King's College London, Leeds, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Khalida Ismail, MBBS, PhD
Department of Psychological Medicine
King's College London
Institute of Psychiatry Psychology and Neuroscience
Denmark Hill Campus, King’s College London, 16 De Crespigny Park
London, SE5 8AB
United Kingdom
Phone: 44 7905985068
Email: khalida.2.ismail@kcl.ac.uk

Abstract

Background: Gestational diabetes mellitus (GDM) is characterized by hyperglycemia in pregnancy and typically resolves after birth. Women with GDM have an increased risk of developing type 2 diabetes mellitus (T2DM) later in life compared to those with normoglycemic pregnancy. While diabetes prevention interventions (DPIs) have been developed to delay or prevent the onset of T2DM, few studies have provided process evaluation (PE) data to assess the mechanisms of impact, quality of implementation, or contextual factors that may influence the effectiveness of the intervention.

Objective: This study aims to identify and evaluate PE data and how these link to outcomes of randomized controlled trials (RCTs) of T2DM prevention interventions for women with GDM.

Methods: A systematic review was conducted to identify studies published from 2005 to 2020 aiming to capture the most recent DPIs. Five electronic bibliographic databases (Cochrane Library, Cochrane Collaboration Registry of Controlled Trials, Embase, PubMed, and MEDLINE) were searched to identify relevant studies. Inclusion criteria were published (peer-reviewed) RCTs of DPIs in women with a current diagnosis or history of GDM. Exclusion criteria were studies not published in English; studies where the target population was women who had a family history of T2D or women who were menopausal or postmenopausal; and gray literature, including abstracts in conference proceedings. The Medical Research Council’s PE framework of complex interventions was used to identify key PE components. The Mixed Method Appraisal Tool was used to assess the quality of included studies.

Results: A total of 24 studies were included; however, only 5 studies explicitly reported a PE theoretical framework. The studies involved 3 methods of intervention delivery, including in person (n=7), digital (n=7), and hybrid (n=9). Two of the studies conducted pilot RCTs assessing the feasibility and acceptability of their interventions, including recruitment, participation, retention, program implementation, adherence, and satisfaction, and 1 study assessed the efficacy of a questionnaire to promote food and vegetable intake. While most studies linked PE data with study outcomes, it was unclear which of the reported PE components were specifically linked to the positive outcomes.

Conclusions: While the Medical Research Council's framework is a valuable source for conducting systematic reviews on PEs, it has been criticized for lacking practical advice on how to conduct them. The lack of information on PE frameworks in our review also made it difficult to categorize individual PE components against the framework. We need clearer guidance and robust frameworks for conducting PEs for the development and reporting of DPIs for women with GDM.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020208212; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=208212

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.1177/16094069211034010

(Interact J Med Res 2025;14:e51718) doi:10.2196/51718

KEYWORDS

gestational diabetes mellitus; randomized controlled trial; process evaluation; implementation; complex interventions

Introduction

Gestational diabetes mellitus (GDM) is characterized by hyperglycemia first recognized in pregnancy and currently impacts between 8% and 23% of pregnancies globally [1]. GDM typically resolves after birth but can have significant implications for both short- and long-term health of women and babies [2]. Women with prior GDM have a nearly 10-fold increased risk of developing type 2 diabetes mellitus (T2DM) than those with normoglycemic pregnancy [3]. Risk factors for T2DM after GDM include high BMI, increasing age, multiparity, poor glucose tolerance, and prepregnancy complications [4-6]. Lifestyle interventions that target both diet and exercise are associated with small but significant effects in reducing the risk of T2DM in women with GDM [7-10].

Process evaluation (PE) theoretical frameworks provide a systematic approach to planning the design of health behavior change interventions [11,12]. However, many of the randomized controlled trials (RCTs) reported not using these PE frameworks to plan the design of lifestyle interventions [11,13-17]. Several key PE theoretical frameworks have been developed and widely used [13-17], and researchers have made progress in updating the methodologies and definitions of key PE components to construct more comprehensive frameworks to measure the success and effectiveness of interventions [13]. Knowledge of these processes can better inform future policy and practice [18] and provide opportunities to improve study design and methodologies of future diabetes prevention interventions (DPIs) [19].

The UK Medical Research Council (MRC) theoretical framework on PEs for complex interventions was chosen for this systematic review [15], as it is well regarded and recognizes the need for more formal guidance on how to conduct PEs. The framework considers processes along 3 interlinked dimensions: implementation, mechanisms of impact, and contextual factors. Table 1 presents a glossary of key PE components and methodologies [15].

The overarching aim of this review was to identify and evaluate any PE conducted and how these link to outcomes of RCTs of DPIs after GDM. Specific objectives were to identify the extent to which PE components were reported and described in RCTs of DPIs for women with previous or current GDM and to assess whether these components could contribute to explaining the intervention outcomes of the DPIs.

Table 1. Glossary of key process evaluation components and methodologies.

Key process evaluation components	Definitions	Quantitative methods	Qualitative methods
Implementation			
Content delivered	The content included in the intervention and the method in which the content is delivered to the participants.	• Structured observation	• Observational study
Method of content delivery	How the intervention content was delivered to the participants and the extent to which quantitative and qualitative methods were used.	• Structured observation	• Observational study
Fidelity	The extent to which the intervention was delivered as intended. Fidelity represents the quality and integrity of the intervention.	• Behavioral coding systems • Questionnaires • Protocol checklist • Structured observation	• Audiotapes of sessions • Observational study
Dosage delivered	The quantity or number of intended units delivered during the intervention including the component delivered and the extent of the participant's engagement.	• Checklist records of dose delivered • Structured observation	• Audiotapes of sessions
Dosage received	The quantity or number of intended units received during the intervention including the component received by the participants.	• Behavioral coding systems • Questionnaires • Structured observation • Digital monitoring (eg, digital feedback)	• Audiotapes of sessions • Focus groups • Interviews • Observational monitoring
Adaptations	The extent to which alterations were made to an intervention to achieve better contextual fit.	• Structured observation	• Observational study
Reach	The proportion of the intended target audience that comes into contact with and participated in the intervention.	• Attendance lists • Standardized protocols	Observational monitoring
Mechanisms of impact			
Participant responses toward the intervention	The responses of participants who interacted with and received the intervention, their satisfaction, and the degree to which they found the intervention acceptable.	• Questionnaires	• Interviews • Focus groups
Mediators	The extent to which intermediary processes inform subsequent changes in outcomes.	• Structured observations	• Observational study
Unexpected pathways or consequences	The extent to which identifying unexpected pathways and mechanisms during the intervention meets the research needs and leads to intervention outcomes.	• Structured observations	• Observational study
Contextual factors			
Barriers	Any external factors that may act as a barrier toward the intervention implementation or its effect on the outcomes.	• Statistical analysis • Questionnaires	• Interviews • Focus groups • Observational study
Facilitators	Any external factors that may act as a facilitator to the intervention implementation or its effect on the outcomes.	• Statistical analysis • Questionnaires	• Interviews • Focus groups • Observational study

Methods

Study Design

This review used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was registered on PROSPERO (CRD42020208212). Owing to limited guidance on conducting systematic reviews in this area, we used the UK MRC theoretical framework from which we derived a glossary of components to use as a checklist to extract and categorize qualitative and quantitative data (Table 1) [15]. This framework considers 3 linked dimensions in which to consider processes: implementation, mechanisms of impact, and contextual factors. Implementation involves an assessment of fidelity, dose delivered and received (ie, how often or effectively it was delivered), adaptations made, reach, content delivered, and method of content delivery [20,21]. Mechanisms of impact involve an assessment of participants' satisfaction with the intervention, "unexpected consequences," and "mediators" [18]. Contextual factors refer to the external barriers and facilitators, such as cultural or organizational factors, that may alter the implementation of an intervention [22].

Search Strategy

The timeframe for the search was limited to between December 1, 2005, and December 16, 2020. This timeframe was selected because of the rapid increase in DPIs in addition to the fast-moving space of digital technology. The review aimed to capture the most recent DPIs. Five electronic bibliographic databases were used: Cochrane Library; Cochrane Collaboration Registry of Controlled Trials; Embase; PubMed; and MEDLINE. The main key search terms were GDM, RCT, and PE (see Multimedia Appendix 1 for a detailed list of search terms and strings and Multimedia Appendix 2 for key definitions). Reference lists of the included studies were also searched for additional eligible studies. Boolean search was used to combine the keywords with operators such as AND, NOT, and OR to further produce more relevant results, for example: ((GDM OR gestational diabetes OR pregnancy-induced diabetes OR diabetes in pregnancy) AND ((RCT OR controlled clinical trial OR pragmatic control trial OR clinical trial) AND (process evaluation OR program evaluation OR process assessment OR process acceptance OR outcome measures)); ((GDMs OR gestational diabetes OR pregnancy-induced diabetes)) AND ((RCT OR controlled clinical trial OR pragmatic clinical trial OR clinical trial) OR (process evaluation OR program assessment OR process acceptance OR outcome measures).

Eligibility Criteria

Inclusion criteria were published (peer-reviewed) RCTs of DPIs in women with a current diagnosis or history of GDM. Exclusion criteria were studies not published in English; studies where the target population was women who had a family history of T2D or women who were menopausal or postmenopausal; and gray literature, including abstracts in conference proceedings. For further details about the eligibility criteria, please see the study protocol [23].

Study Selection

After initial deduplication of the extracted data using Endnote (Clarivate), studies were uploaded to the platform Rayyan for screening. The 2 reviewers (IIMS and MB) independently screened the titles and abstracts of the studies. Disagreements were resolved via consensus or decided by a third reviewer (IPN) when consensus could not be reached. The full texts of potentially relevant studies were retrieved for further screening and independently appraised by 2 reviewers (IIMS and MB) for final inclusion.

Data Extraction, Study Characteristics, and Analysis

Standardized data extraction forms were developed and piloted. Extracted data included a summary of study characteristics, evaluation of processes (if any), methods, and findings. Data were extracted by a single reviewer (IIMS) and verified by a second reviewer (NH); discrepancies were again resolved via consensus.

We conducted a narrative synthesis of the studies as we anticipated that the heterogeneity of the methods used to assess processes would preclude meta-analysis, as has been found previously [10].

Assessment of Quality

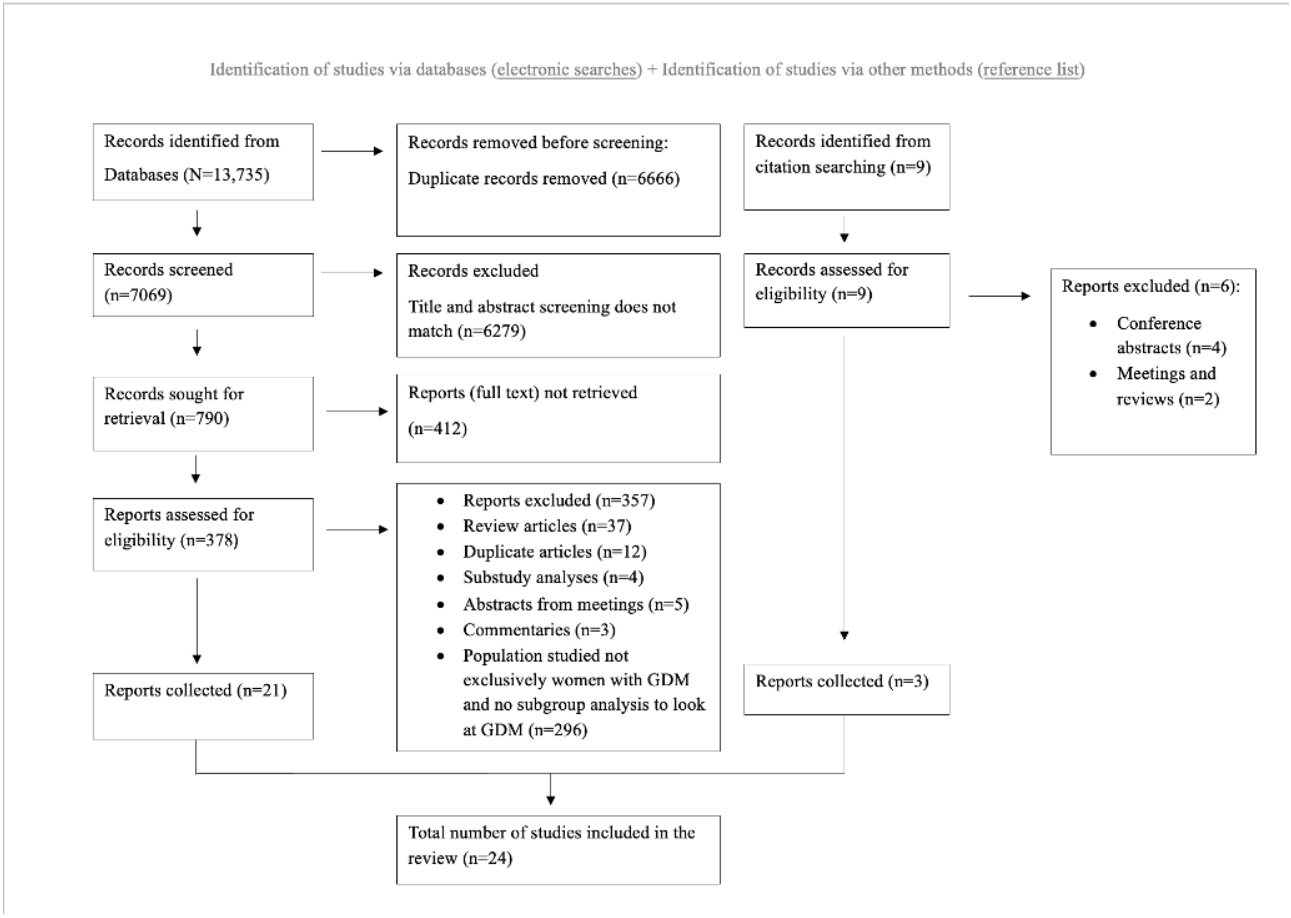
The Mixed Method Appraisal Tool was used to assess the quality of included studies by 2 independent reviewers (IIMS) and (NH). The discrepancies were again resolved via consensus.

Results

Study Selection

The PRISMA flow diagram for this review is shown in Figure 1. Of the 13,735 records initially identified, the full text of 378 full-text studies were screened for eligibility and 21 studies met the eligibility criteria. An additional 3 studies were added following citation searching. Therefore, 24 studies were included in this review. Most of these studies fulfilled at least 3 methodological quality criteria outlined by the Mixed Method Appraisal Tool for each study design (Multimedia Appendix 3 [3,24-46]).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of search. GDM: gestational diabetes mellitus.



Study Characteristics

The stated primary aims of all 24 DPI studies reviewed were to improve dietary and physical activity outcomes, increase self-efficacy levels and risk perception of T2DM, and decrease maternal postpartum weight and BMI. The 24 studies evaluated in-person (n=8), digital (n=7), and hybrid (n=9) interventions. Two of the studies were pilot studies of the feasibility and acceptability of DPIs. All studies were published between 2007 and 2020. Most studies (n=22, 92%) were conducted in high-income countries. Only 2 studies were conducted in low- to middle-income countries: Egypt (n=1) and Malaysia (n=1). The sample size ranged from 31 to 1180 participants. The intervention duration varied from 12 weeks to 6 years. See Multimedia Appendix 4 [3,24-46] for a description of study characteristics of the studies included in this review.

Number and Type of PE Components Reported

The most frequent PE components evaluated were content delivered and method of content delivery (n=24), dose delivered (n=20), reach (n=20), dose received (n=17), barriers (n=15), facilitators (n=8), and participants' responses toward the intervention (n=8). Relatively few studies reported process data relating to adaptations (n=7), mediators (n=7), fidelity (n=5), or unexpected pathways or consequences (n=2). All PE components outlined by the MRC framework and reported in the reviewed DPIs are summarized in Multimedia Appendices 5 and 6 [3,24-46].

PE Frameworks and Measures Reported

Five studies explicitly referred to a PE framework or PE quantitative or qualitative measures. Only 1 of these referred to process measures in the context of the MRC framework of complex interventions [24]. The others reported measures to ensure adherence to the study protocol [25], to measure "reach" (using the penetration, implementation, participation, and effectiveness metric) [26,27], or to assess participants' satisfaction to the program using a survey [47].

Implementation

Content Delivered and Method of Content Delivery

Eight studies evaluated in-person or group DPIs [3,28-34], 7 evaluated digital DPIs [35-40,48], and 9 delivered hybrid DPIs [24-27,41-44]. Most of the interventions were multicomponent that focused on diet and self-directed physical activity [24-27,33,35,40-46,49] or were facilitator led [29,30,43]. One intervention also included advice on breastfeeding [28].

Behavior Techniques and Theoretical Frameworks Used to Deliver Content

Behavior change techniques included motivational interviewing (MI) with individual goal setting [26,29]; enhancing healthy lifestyle change and self-management [24]; guided self-help [33]; nutritional and physical activity recommendations [30]; self-help guidance based on responses to a dietary food intake questionnaire [3]; and guideline-based nutritional strategies [44]. Five interventions were underpinned by established

theoretical frameworks—social cognitive theory and the transtheoretical model of behavior change [25]; the Health Action Process Approach supported by social cognitive and self-regulation theory [27] based on a Health Belief Model [50]; and messages on weight loss, diet, and physical activity based on a harm reduction model to promote healthy lifestyle changes [43].

The link between a process and outcome was reported in 2 studies. In the first, the use of behavior change strategies was reported to improve women's knowledge, beliefs, and self-reported practices as well as decrease postpartum weight gain in 1 study [33]. In the second study, there were no observed changes in weight, waist circumference, or blood glucose level and this was correlated with no changes in health perception or self-efficacy [3].

Fidelity

Five DPIs (digital: n=3 and hybrid: n=2) reported on the fidelity (quality) of intervention implementation [25,27,36,39,45]. Assessment of fidelity involved recording telephone calls or sessions [36,45], a qualitative [45] or quantitative [36,39] assessment of adherence, and sometimes also involved input from the participants [45]. Two of the studies attributed a high level of fidelity to effective weight loss and weight goals [27,36].

Dose Delivered

Nineteen studies reported on the quantity, frequency, and duration of the interventions [24-26,28-30,32,34,36,39-43,48]. Most studies delivered around 1 to 4 sessions lasting 2 hours. The shortest dose duration was 12 weeks and included 12-weekly sessions of 2.5 hours including group physical activity, health education, and an individualized lifestyle counseling session [29].

Further details in relation to dose delivery were generally limited. For instance, the frequency [24,40] or quantity of sessions or emails [43,45] were not reported. Similarly, 3 digital DPIs of web-based interventions involving health websites [40], internet telemedicine [37], or pedometer messaging [38] did not report the quantity of health content and messages sent by health care professionals. Borgen et al [35] reported the delivery of their mobile health app but did not mention how long it would take to peruse the app's content. Finally, another web-based intervention [44] did not report on how frequently participants should use the pedometer or step count goal target.

Dose Received

Sixteen of the studies reported data on the dose received by the participants [3,24-26,28-30,32,36-38,43,45]. The dose response included the average number of sessions attended by women [25-27,29,36,40,43-45]; those returning documents or questionnaires [51-53]; combined food intake and physical activity logbooks [38,54,55]; and log-ins to websites or systems [37,40,45].

Information on the dose received was used to highlight the willingness of participants to take part in the study and the subsequent success of the intervention. For example, in 1 hybrid study [44], it was reported that most women in the study attended all 4 lifestyle counseling sessions in their intervention, concluding that it was successful in reducing weight and increasing physical activity levels. Details on the dose received were also used to explain any lack of intervention effects. In another hybrid study [43], authors attributed their unsuccessful intervention outcomes to the limited number of women logging onto their website intervention and observed that engagement did not improve with fortnightly automated email reminders. One digital study observed that they did not collect data on user engagement for its mobile health app to protect women's privacy and questioned that this may have explained the nonsignificant effect of the intervention [35].

Adaptations

Adaptations were described in 7 DPIs, including 2 in-person [38,55], 2 digital [35,45], and 3 hybrid DPIs [24,25,32]. These adaptations were seen as important components for the success of the intervention and included providing a choice of intervention methods [28] and tailoring interventions to the characteristics of the target population [45] to specific ethnic or cultural backgrounds [24,34,35]. In 3 hybrid DPIs, MI techniques were used to tailor sessions for participants [24,25] and illustrations and simple messages were added to address low health literacy [41].

Reach

Twenty studies reported on the reach of its intervention in relation to the target population [24-26,28-32,34,36-41,43,45]. Loss to follow-up was a recognized challenge, perhaps reflecting the case mix of the target population. Two of the studies referred to pregnancy complications as a reason for dropout [31,33] with another attributing loss to follow-up to work or personal commitments, initiation of a weight loss diet, subsequent pregnancy, not being contactable, and finding intervention resources unhelpful [40]. One digital DPI reported a low participation rate of 17%, making inference to the broader target population difficult [39]; another reported that few women who received information about their intervention proceeded to enroll for participation [38]. Other studies that used PE methods to measure reach reported low rates [24,27,43] and 2 additional studies had 33% to 39% attrition rates [29,34].

Mechanisms of Impact

Participant Responses Toward the Intervention

Eight studies reported on participant feedback and responses toward the interventions including 4 digital [35,38-40], 3 hybrid [25,26,44], and 1 in-person DPI [29]. Overall, women expressed satisfaction with the interventions [25,29,38,39,44], reporting increased confidence in setting health goals [39,40] and more engagement with health management [35].

Some studies did not receive a positive response. Only a third of women reported being motivated by the website content in 1 digital intervention [40] and even fewer (22% and 31%) had a positive reaction to the text messaging component. Women also criticized the lack of information on optimal carbohydrates to consume when transitioning between pregnancy to postpartum diets and reported the need for more low-fat recipes [25,40].

Mediators

Mediators were described in the interventions in 2 in-person DPI studies [29,34], 4 digital DPIs [35,37,39,40], and 1 hybrid DPI [43], although not explicitly. O'Dea et al [29] found that women valued individual face-to-face sessions with health care professionals when setting health goals which improved their stress levels, diet self-efficacy, and quality of life. Social relationships also developed between the women due to regular attendance at the same lifestyle counseling groups [34]. Digital DPIs were also reported to increase women's engagement in their own health [35], enhance self-efficacy and confidence [37,40], and provide reassurance in developing and attaining health goals [39]. Further, 1 hybrid study [43] qualitatively linked the inclusion of partners in their intervention as a key mediator for women making them more likely to participate in and engage with the intervention.

Unexpected Pathways or Consequences

Only 2 studies [39,41] reported unexpected consequences in relation to their interventions. Carolan-Olah and Sayakhot [41] observed there was an unusually high percentage of women in the intervention group who had attended their postpartum oral glucose tolerance test (OGTT) appointment. Although not directly linked to the study, the intervention highlighted the importance of postpartum testing, which may have increased women's motivation to attend future follow-up appointments. Another digital study [39] experienced unexpected challenges with data retrieval from women and general practitioners, resulting in missing baseline blood tests and self-reported data at follow-up. However, this missing data did not appear to negatively influence the findings as women in the intervention group were found to significantly reduce their fat intake.

Contextual Factors

Barriers

Fifteen studies referred to barriers that women faced when participating in interventions [3,24-26,29,31,33-35,37-40,43,44]. Such barriers included lack of time due to family and work commitments, subsequent pregnancy, and changes to daily life due to the demands of motherhood [24,29,31,33,40,44] technology issues, and lack of internet access [24,38,43]. Similarly, unavailability due to childcare responsibilities and forgetting to attend study visits prevented mothers from engaging in in-person lifestyle counseling sessions in the nondigital interventions [3,24,29]. Another in-person DPI acknowledged how culture plays a role in hindering women from participating in and completing an intervention when they are placed under pressure to commit to family and social norms [34].

Facilitators

The 8 DPI studies including 4 in-person DPIs [28,29,31,34], 2 hybrid DPIs [43,44], and 2 digital DPIs [37,40] described facilitators as contextual factors which may have influenced the outcomes of their interventions. The most common facilitator identified in 1 hybrid and 1 in-person DPI was receiving support from a partner during the intervention [29,43] and having higher levels of income and personal education background [43]. Having access to healthy food, food vouchers, a babysitter, and an exercise buddy were also reported as facilitators to assist and support the uptake and maintenance of postpartum healthy behaviors [31,40]. The competency and skills of health care professionals delivering the interventions were acknowledged as facilitators to ensure the uptake of the intervention [28]. Having an accessible, central location for assessments and employers who allowed time for visits also facilitated participation [34,44].

Discussion

Overview

We identified and reviewed 24 RCTS of DPIs that evaluated at least 1 PE component for women with previous or current GDM. Only 10 studies explicitly reported individual processes, and only 5 explicitly referred to a PE framework or PE quantitative or qualitative measures to assess processes. Overall, few studies reported and evaluated processes in relation to study outcomes, and although most DPIs linked PE components and DPI outcomes in some capacity, it was challenging to attribute any 1 process component to the effectiveness and success of the intervention.

Summary of Key Findings

The most complete of the PE components reported across all 24 studies was the combined component of content delivered and method of content delivery. Clear narratives were provided on how the interventions were delivered and their content. Most of the interventions also referred to broad behavior change techniques such as goal setting and MI, which are known to be effective in promoting lifestyle behavior change [56,57]. However, few studies linked these theories with mechanisms of action by providing a detailed psychological or behavior change theoretical framework on which their intervention was underpinned. Such frameworks are important to make sense of "how," "why," and "under what circumstances" intervention components work together to achieve the desired outcomes. Moving forward, we recommend that all RCTs of DPIs publish an overview of the theories they have used to design and evaluate the intervention for greater transparency.

Although most DPIs reported information on the dose delivered, PE measures on the dose received were less complete. Generally, increased frequency and longer sessions resulted in more engagement with the intervention and better outcomes [26-28,30,31,36]. More frequent sessions may also be preferred by the participants: for a study on the use of MI and hemoglobin A_{1c} outcomes, women said that they would like more sessions to improve their engagement with the intervention [58]. The timing of the intervention is also a consideration. Most of the DPIs that we reviewed delivered interventions during either the antenatal or postpartum periods. However, systematic review evidence suggests that optimal pregnancy outcomes can be achieved by delivering the intervention during both periods since engagement with a healthy lifestyle poses tends to be poorer during the postpartum stage [5,51]. Moreover, there was little information in the studies reviewed on how the dose received impacted the success of the intervention. The challenges in assessing adherence to the dose received are well

recognized as people may have received a dose but choose not to take it [59]. This is also supported in the literature where a nurse-led psychological intervention for a T2D cluster RCT study identified a lower dose received than intended, but that this was not associated with the dose delivered [60]. Further research to explore how timing, frequency, and duration of the intervention affect outcomes among women with current or previous GDM is indicated.

Although all 24 studies included in this review sought to describe the reach of the intervention, few gave a detailed evaluation of this measure. Poor reach was attributed to generic reasons, most commonly low participation rates, loss to follow-up, and small sample sizes [26,27,29,36,39]. The increase in OGTT appointment attendance found in 1 study [41] is contrary to prior research linking loss to follow-up and withdrawal numbers to participants' reluctance to attend OGTT appointments [49,54], which highlights the influence that different contexts have on outcomes in this population. Moreover, the studies that made adaptations to tailor their DPIs to personal or cultural needs [24,25,28,34,35,41,45] are likely to have increased reach by making the interventions more relevant. A clear definition of the target population and consideration of how the DPI can be adapted to make it as inclusive as possible seems important for its success.

Most of the studies reported on barriers to intervention engagement. Similar to previous systematic reviews [10,61], the barriers primarily related to lack of time, childcare commitments, and challenges to maintaining healthy lifestyles [24,29]. Focus groups with women suggest that these inherent barriers can be mitigated by combining direct contact with health care professionals with web-based interventions [48]. Moreover, it is worth noting that, despite the many barriers to participation identified, women were still in favor of the DPIs, and many resulted in positive outcomes. This aligns with previous research in T2D [62] and supports a need and willingness among women with previous or current GDM to engage with DPIs.

A particular concern identified by this review was the lack of reporting or evaluation of fidelity. Only 5 studies reported on this process component. Ensuring fidelity is a key aspect of any DPI because it safeguards against nonadherence to the study protocols and inadequate implementation delivery [47,63]. Ideally, fidelity should be measured prior to conducting full-scale implementation to distinguish between outcomes that are related to ineffectiveness from those related to protocol deviation [64,65]. In our review, methods to assess fidelity included audiotaping and patient registers of attendance [27,36]. Such methods have been tried and tested in the T2D setting [47,63] and are likely to be more robust than relying on the expertise of health care professionals as a fidelity measure [66]. To minimize research waste, fidelity assessments should be commenced at the design stage of a trial and incorporate standardized measures where possible.

Finally, we welcome the use of both quantitative and qualitative methods to evaluate the PE components in our studies because this enables a more in-depth understanding of the relationship between individual components and outcomes [52,67]. We reemphasize the importance of mixed methods when conducting thorough PE.

Strengths and Limitations

To the best of our knowledge, this is the first and only systematic review of PEs of T2DM prevention interventions for women with GDM. We used robust methodology, an established framework of PE for complex interventions [15], and restricted the review to RCTs, which are known to be the best design for evaluating complex health care interventions [15]. The glossary list of standardized DPI content and PE terms developed during the review process can be used to guide future RCTs in this area. However, while the MRC framework [18] is a valuable source for conducting systematic reviews on PEs, it has been criticized for lacking practical advice on how to conduct them [68]. The lack of information on PE frameworks in our review also made it difficult to categorize individual PE components against the framework. Moreover, there is a lack of consensus on how PE components should be defined and interpreted by researchers, which is only partially addressed in the MRC guidelines [15], which facilitate understanding of PE constructs but provide little clarity on what should be measured and how.

Conclusions

This systematic review has highlighted that there are important gaps in the reporting of PE metrics for RCTs of T2DM in women with GDM. We recommend rigorous, systematic, and in-depth PE guidance to facilitate reporting of these studies. Future research should focus on reaching consensus on the reporting of PE measures using established frameworks and evaluating PE in real-world health care settings to optimize the interpretation of study outcomes.

Acknowledgments

This work is supported by MY PAIR Grant UK-Malaysia: Joint Partnership Call on Non-Communicable Diseases (Malaysia: JPT.S(BPKI)2000/011/06/05 (27); United Kingdom: MR/T018240/1). This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London and Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

The authors confirm that generative artificial intelligence was not used to conduct or report any portion of this systematic review.

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Authors' Contributions

IIMS developed the methodology, conducted the investigation and formal analysis, and cowrote and reviewed the original draft. MB conducted the investigation, cowrote the original draft, reviewed and edited subsequent drafts, supervised junior researchers, and contributed to project administration. IPN codeveloped the original concept, methodology, and investigation; curated data and cowrote the original draft; edited subsequent drafts; supervised junior researchers; and acquired funding. AB cowrote the original draft and reviewed subsequent drafts. MP developed the original methodology, contributed to the investigation, cowrote drafts, and supervised junior researchers. BNMY contributed to the writing of drafts and reviewing subsequent versions. SB contributed to writing, reviewed, and edited subsequent versions. CSM codeveloped the original concept, contributed to the writing and reviewing of subsequent versions, and supervised junior researchers. AF codeveloped the original concept, developed the methodology, cowrote and reviewed subsequent versions of the original draft, and supervised junior researchers. KI codeveloped the original concept; developed the methodology; conducted the formal analysis; contributed to the investigation, data curation, and inputting of data into the software; supervised junior researchers; contributed to project administration; and acquired funding. BHC codeveloped the original concept, developed the methodology, cowrote the original draft and reviewed subsequent versions, supervised junior researchers, and acquired funding. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed list of search terms and strings.
[DOCX File , 16 KB - ijmr_v14i1e51718_app1.docx]

Multimedia Appendix 2
Key definitions of types and formats of DPI content delivered, behavior change techniques, and psychological theoretical frameworks embedded within DPIs. DPI: diabetes prevention intervention.
[DOCX File , 17 KB - ijmr_v14i1e51718_app2.docx]

Multimedia Appendix 3
Quality assessment of review articles.
[DOCX File , 26 KB - ijmr_v14i1e51718_app3.docx]

Multimedia Appendix 4
Study characteristics of interventions for the prevention of type 2 diabetes in women with gestational diabetes mellitus.
[DOCX File , 47 KB - ijmr_v14i1e51718_app4.docx]

Multimedia Appendix 5
Process evaluation implementation components for T2DM prevention lifestyle intervention studies on women with GDM. GDM: gestational diabetes mellitus; T2DM: type 2 diabetes mellitus.
[DOCX File , 31 KB - ijmr_v14i1e51718_app5.docx]

Multimedia Appendix 6 [DOCX File , 21 KB - ijmr_v14i1e51718_app6.docx]

Multimedia Appendix 7
PRISMA Checklist for systematic review.
[PDF File (Adobe PDF File), 58 KB - ijmr_v14i1e51718_app7.pdf]

References

1. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
2. Kytö M, Koivusalo S, Tuomonen H, Strömberg L, Ruonala A, Marttinen P, et al. Supporting the management of gestational diabetes mellitus with comprehensive self-tracking: mixed methods study of wearable sensors. JMIR Diabetes 2023 Oct 31;8:e43979 [FREE Full text] [doi: 10.2196/43979] [Medline: 37906216]

3. Vézina-Im LA, Perron J, Lemieux S, Robitaille J. Promoting fruit and vegetable intake in childbearing age women at risk for gestational diabetes mellitus: A randomised controlled trial. J Health Psychol 2019 Apr;24(5):600-612. [doi: 10.1177/1359105316680021] [Medline: 27872389]
4. Baptiste-Roberts K, Barone BB, Gary TL, Golden SH, Wilson LM, Bass EB, et al. Risk factors for type 2 diabetes among women with gestational diabetes: a systematic review. Am J Med 2009 Mar;122(3):207-214.e4 [FREE Full text] [doi: 10.1016/j.amjmed.2008.09.034] [Medline: 19272478]
5. Li Z, Cheng Y, Wang D, Chen H, Chen H, Ming WK, et al. Incidence rate of type 2 diabetes mellitus after gestational diabetes mellitus: a systematic review and meta-analysis of 170,139 women. J Diabetes Res 2020;2020:3076463 [FREE Full text] [doi: 10.1155/2020/3076463] [Medline: 32405502]
6. Rayanagoudar G, Hashi AA, Zamora J, Khan KS, Hitman GA, Thangaratinam S. Quantification of the type 2 diabetes risk in women with gestational diabetes: a systematic review and meta-analysis of 95,750 women. Diabetologia 2016;59(7):1403-1411 [FREE Full text] [doi: 10.1007/s00125-016-3927-2] [Medline: 27073002]
7. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: an exposure-wide umbrella review of meta-analyses. PLoS One 2018;13(3):e0194127 [FREE Full text] [doi: 10.1371/journal.pone.0194127] [Medline: 29558518]
8. Brown J, Alwan NA, West J, Brown S, McKinlay CJ, Farrar D, et al. Lifestyle interventions for the treatment of women with gestational diabetes. Cochrane Database Syst Rev 2017;5(5):CD011970 [FREE Full text] [doi: 10.1002/14651858.CD011970.pub2] [Medline: 28472859]
9. Gilinsky AS, Kirk AF, Hughes AR, Lindsay RS. Lifestyle interventions for type 2 diabetes prevention in women with prior gestational diabetes: a systematic review and meta-analysis of behavioural, anthropometric and metabolic outcomes. Prev Med Rep 2015;2:448-461 [FREE Full text] [doi: 10.1016/j.pmedr.2015.05.009] [Medline: 26844102]
10. Guo J, Chen JL, Whittemore R, Whitaker E. Postpartum lifestyle interventions to prevent type 2 diabetes among women with history of gestational diabetes: a systematic review of randomized clinical trials. J Womens Health 2016;25(1):38-49. [doi: 10.1089/jwh.2015.5262] [Medline: 26700931]
11. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
12. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
13. Baranowski T, Stables G. Process evaluations of the 5-a-day projects. Health Educ Behav 2000;27(2):157-166. [doi: 10.1177/109019810002700202] [Medline: 10768797]
14. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials 2013;14:15 [FREE Full text] [doi: 10.1186/1745-6215-14-15] [Medline: 23311722]
15. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. BMJ 2015;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
16. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005;6(2):134-147. [doi: 10.1177/1524839904273387] [Medline: 15855283]
17. Steckler AE, Linnan LE. Process Evaluation for Public Health Interventions and Research. San Francisco, CA: Jossey-Bass; 2002.
18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new medical research council guidance. BMJ 2008;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
19. Whittemore R, Grey M. The systematic development of nursing interventions. J Nurs Scholarsh 2002;34(2):115-120. [doi: 10.1111/j.1547-5069.2002.00115.x] [Medline: 12078534]
20. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007;2:40 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
21. Montgomery P, Underhill K, Gardner F, Operario D, Mayo-Wilson E. The Oxford Implementation Index: a new tool for incorporating implementation data into systematic reviews and meta-analyses. J Clin Epidemiol 2013;66(8):874-882 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.03.006] [Medline: 23810026]
22. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008;41:327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]
23. Mohd Sa'id II, Papachristou Nadal I, Forbes A, Goldsmith K, Ismail IZ, Hassan F, et al. A protocol of process evaluations of interventions for the prevention of type 2 diabetes in women with gestational diabetes mellitus: a systematic review. Int J Qual Methods 2021;20:160940692110340. [doi: 10.1177/16094069211034010]

24. Holmes VA, Draffin CR, Patterson CC, Francis L, Irwin J, McConnell M, et al. Postnatal lifestyle intervention for overweight women with previous gestational diabetes: a randomized controlled trial. J Clin Endocrinol Metab 2018;103(7):2478-2487. [doi: 10.1210/jc.2017-02654] [Medline: 29762737]
25. Ferrara A, Hedderson MM, Albright CL, Ehrlich SF, Quesenberry CP, Peng T, et al. A pregnancy and postpartum lifestyle intervention in women with gestational diabetes mellitus reduces diabetes risk factors: a feasibility randomized control trial. Diabetes Care 2011;34(7):1519-1525 [FREE Full text] [doi: 10.2337/dc10-2221] [Medline: 21540430]
26. Lipscombe LL, Delos-Reyes F, Glenn AJ, de Sequeira S, Liang X, Grant S, et al. The avoiding diabetes after pregnancy trial in moms program: feasibility of a diabetes prevention program for women with recent gestational diabetes mellitus. Can J Diabetes 2019;43(8):613-620. [doi: 10.1016/j.jcjd.2019.08.019] [Medline: 31669188]
27. O'Reilly SL, Dunbar JA, Versace V, Janus E, Best JD, Carter R, et al. Mothers After Gestational Diabetes in Australia (MAGDA): a randomised controlled trial of a postnatal diabetes prevention program. PLoS Med 2016;13(7):e1002092 [FREE Full text] [doi: 10.1371/journal.pmed.1002092] [Medline: 27459502]
28. Koivusalo SB, Rönö K, Klemetti MK, Roine RP, Lindström J, Erkkola M, et al. Gestational diabetes mellitus can be prevented by lifestyle intervention: the Finnish Gestational Diabetes Prevention Study (RADIEL): a randomised controlled trial. Diabetes Care 2016;39(1):24-30. [doi: 10.2337/dc15-0511] [Medline: 26223239]
29. O'Dea A, Tierney M, McGuire BE, Newell J, Glynn LG, Gibson I, et al. Can the onset of type 2 diabetes be delayed by a group-based lifestyle intervention in women with prediabetes following gestational diabetes mellitus (GDM)? Findings from a randomized control mixed methods trial. J Diabetes Res 2015;2015:798460 [FREE Full text] [doi: 10.1155/2015/798460] [Medline: 26347894]
30. Pérez-Ferre N, Del Valle L, Torrejón MJ, Barca I, Calvo MI, Matía P, et al. Diabetes mellitus and abnormal glucose tolerance development after gestational diabetes: a three-year, prospective, randomized, clinical-based, mediterranean lifestyle interventional study with parallel groups. Clin Nutr 2015;34(4):579-585. [doi: 10.1016/j.clnu.2014.09.005] [Medline: 25262459]
31. Shek NWM, Ngai CSW, Lee CP, Chan JYC, Lao TTH. Lifestyle modifications in the development of diabetes mellitus and metabolic syndrome in Chinese women who had gestational diabetes mellitus: a randomized interventional trial. Arch Gynecol Obstet 2014;289(2):319-327. [doi: 10.1007/s00404-013-2971-0] [Medline: 23897066]
32. Shyam S, Arshad F, Abdul Ghani R, Wahab NA, Safii NS, Nisak MYB, et al. Low glycaemic index diets improve glucose tolerance and body weight in women with previous history of gestational diabetes: a six months randomized trial. Nutr J 2013;12(1):1-12 [FREE Full text] [doi: 10.1186/1475-2891-12-68] [Medline: 23705645]
33. Tawfik MY. The impact of health education intervention for prevention and early detection of type 2 diabetes in women with gestational diabetes. J Community Health 2017;42(3):500-510. [doi: 10.1007/s10900-016-0282-7] [Medline: 27743337]
34. Zilberman-Kravits D, Meyerstein N, Abu-Rabia Y, Wiznitzer A, Harman-Boehm I. The impact of a cultural lifestyle intervention on metabolic parameters after gestational diabetes mellitus a randomized controlled trial. Matern Child Health J 2018;22(6):803-811. [doi: 10.1007/s10995-018-2450-0] [Medline: 29411251]
35. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in norway. BMJ Open 2019;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
36. Ferrara A, Hedderson MM, Brown SD, Albright CL, Ehrlich SF, Tsai AL, et al. The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster randomized controlled trial. Diabetes Care 2016;39(1):65-74 [FREE Full text] [doi: 10.2337/dc15-1254] [Medline: 26657945]
37. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
38. Kim C, Draska M, Hess ML, Wilson EJ, Richardson CR. A web-based pedometer programme in women with a recent history of gestational diabetes. Diabet Med 2012;29(2):278-283 [FREE Full text] [doi: 10.1111/j.1464-5491.2011.03415.x] [Medline: 21838764]
39. Reinhardt JA, van der Ploeg HP, Grzegrzulka R, Timperley JG. lmplementing lifestyle change through phone-based motivational interviewing in rural-based women with previous gestational diabetes mellitus. Health Promot J Austr 2012;23(1):5-9. [doi: 10.1071/he12005] [Medline: 22730940]
40. Rollo ME, Baldwin JN, Hutchesson M, Aguiar EJ, Wynne K, Young A, et al. The feasibility and preliminary efficacy of an eHealth lifestyle program in women with recent gestational diabetes mellitus: a pilot study. Int J Environ Res Public Health 2020;17(19):7115. [doi: 10.3390/ijerph17197115] [Medline: 32998401]
41. Carolan-Olah M, Sayakhot P. A randomized controlled trial of a web-based education intervention for women with gestational diabetes mellitus. Midwifery 2019;68:39-47. [doi: 10.1016/j.midw.2018.08.019] [Medline: 30343264]
42. Hu G, Tian H, Zhang F, Liu H, Zhang C, Zhang S, et al. Tianjin Gestational Diabetes Mellitus Prevention Program: study design, methods, and 1-year interim report on the feasibility of lifestyle intervention program. Diabetes Res Clin Pract 2012;98(3):508-517. [doi: 10.1016/j.diabres.2012.09.015] [Medline: 23010556]

43. McManus R, Miller D, Mottola M, Giroux I, Donovan L. Translating healthy living messages to postpartum women and their partners after gestational diabetes (GDM): body habitus, A1C, lifestyle habits, and program engagement results from the families defeating diabetes (FDD) randomized trial. Am J Health Promot 2018;32(6):1438-1446. [doi: 10.1177/0890117117738210] [Medline: 29108443]
44. Peacock AS, Bogossian FE, Wilkinson SA, Gibbons KS, Kim C, McIntyre HD. A randomised controlled trial to delay or prevent type 2 diabetes after gestational diabetes: walking for exercise and nutrition to prevent diabetes for you. Int J Endocrinol 2015;2015:423717 [FREE Full text] [doi: 10.1155/2015/423717] [Medline: 26089886]
45. Nicklas JM, Zera CA, England LJ, Rosner BA, Horton E, Levkoff SE, et al. A web-based lifestyle intervention for women with recent gestational diabetes mellitus: a randomized controlled trial. Obstet Gynecol 2014;124(3):563-570 [FREE Full text] [doi: 10.1097/AOG.0000000000000420] [Medline: 25162257]
46. Liu H, Wang L, Zhang S, Leng J, Li N, Li W, et al. One-year weight losses in the Tianjin Gestational Diabetes Mellitus Prevention Programme: a randomized clinical trial. Diabetes Obes Metab 2018;20(5):1246-1255 [FREE Full text] [doi: 10.1111/dom.13225] [Medline: 29360237]
47. West DS, DiLillo V, Bursac Z, Gore SA, Greene PG. Motivational interviewing improves weight loss in women with type 2 diabetes. Diabetes Care 2007;30(5):1081-1087. [doi: 10.2337/dc06-1966] [Medline: 17337504]
48. Nicklas JM, Zera CA, Seely EW, Abdul-Rahim ZS, Rudloff ND, Levkoff SE. Identifying postpartum intervention approaches to prevent type 2 diabetes in women with a history of gestational diabetes. BMC Pregnancy Childbirth 2011;11:23 [FREE Full text] [doi: 10.1186/1471-2393-11-23] [Medline: 21435246]
49. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002;25(10):1862-1868. [doi: 10.2337/diacare.25.10.1862] [Medline: 12351492]
50. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
51. Huang S, Magny-Normilus C, McMahon E, Whittemore R. Systematic review of lifestyle interventions for gestational diabetes mellitus in pregnancy and the postpartum period. J Obstet Gynecol Neonatal Nurs 2022;51(2):115-125 [FREE Full text] [doi: 10.1016/j.jogn.2021.10.007] [Medline: 34843670]
52. O'Cathain A, Murphy E, Nicholl J. Why, and how, mixed methods research is undertaken in health services research in England: a mixed methods study. BMC Health Serv Res 2007;7:85 [FREE Full text] [doi: 10.1186/1472-6963-7-85] [Medline: 17570838]
53. Gollwitzer PM, Oettingen G. Implementation intentions. In: Encyclopedia of Behavioural Medicine. Cham, Switzerland: Springer International Publishing; 2020:1159-1164.
54. Buckley BS, Harreiter J, Damm P, Corcoy R, Chico A, Simmons D, et al. Gestational diabetes mellitus in Europe: prevalence, current screening practice and barriers to screening. A review. Diabet Med 2012;29(7):844-854. [doi: 10.1111/j.1464-5491.2011.03541.x] [Medline: 22150506]
55. Schunk DH, Usher EL. Social cognitive theory and motivation. In: Ryan RM, editor. The Oxford Handbook of Human Motivation. New York, NY: Oxford University Press; 2012:13-27.
56. Michie S. Designing and implementing behaviour change interventions to improve population health. J Health Serv Res Policy 2008;13:64-69. [doi: 10.1258/jhsrp.2008.008014] [Medline: 18806194]
57. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
58. Waker CL. Effects of motivational interviewing on diabetes self-management behaviors and glycemic control in type 2 diabetes: a translational study [dissertation]. OH: University of Cincinnati; 2012. URL: http://rave.ohiolink.edu/etdc/view?acc_num=ucin1342728810 [accessed 2024-08-26]
59. Graves H, Ismail K, Winkley K. A framework for process evaluations of psychological interventions [Dissertation]. King's College London. 2020. URL: https://kclpure.kcl.ac.uk/portal/en/studentTheses/a-framework-for-process-evaluations-of-psychological-intervention [accessed 2024-08-26]
60. Ismail K, Winkley K, de Zoysa N, Patel A, Heslin M, Graves H, et al. Nurse-led psychological intervention for type 2 diabetes: a cluster randomised controlled trial (Diabetes-6 study) in primary care. Br J Gen Pract 2018;68(673):e531-e540 [FREE Full text] [doi: 10.3399/bjgp18X696185] [Medline: 30012812]
61. Chasan-Taber L. Lifestyle interventions to reduce risk of diabetes among women with prior gestational diabetes mellitus. Best Pract Res Clin Obstet Gynaecol 2015;29(1):110-122 [FREE Full text] [doi: 10.1016/j.bpobgyn.2014.04.019] [Medline: 25220104]
62. Dennison RA, Ward RJ, Griffin SJ, Usher-Smith JA. Women's views on lifestyle changes to reduce the risk of developing type 2 diabetes after gestational diabetes: a systematic review, qualitative synthesis and recommendations for practice. Diabet Med 2019;36(6):702-717 [FREE Full text] [doi: 10.1111/dme.13926] [Medline: 30723968]
63. Campbell EM, Redman S, Moffitt PS, Sanson-Fisher RW. The relative effectiveness of educational and behavioral instruction programs for patients with NIDDM: a randomized trial. Diabetes Educ 1996;22(4):379-386. [doi: 10.1177/014572179602200412] [Medline: 8846745]

64. Bumbarger B, Perkins D. After randomised trials: issues related to dissemination of evidence - based interventions. J Child Serv 2008;3(2):55-64. [doi: 10.1108/17466660200800012]
65. Hawe P, Shiell A, Riley T. Complex interventions: How "out of control" can a randomised controlled trial be? BMJ 2004;328(7455):1561-1563 [FREE Full text] [doi: 10.1136/bmj.328.7455.1561] [Medline: 15217878]
66. Gabbay RA, Añel-Tiangco RM, Dellasega C, Mauger DT, Adelman A, van Horn DHA. Diabetes nurse case management and motivational interviewing for change (DYNAMIC): results of a 2-year randomized controlled pragmatic trial. J Diabetes 2013;5(3):349-357 [FREE Full text] [doi: 10.1111/1753-0407.12030] [Medline: 23368423]
67. Curry LA, Nembhard IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. Circulation 2009;119(10):1442-1452. [doi: 10.1161/CIRCULATIONAHA.107.742775] [Medline: 19289649]
68. Evans R, Scourfield J, Murphy S. Pragmatic, formative process evaluations of complex interventions and why we need more of them. J Epidemiol Community Health 2015;69(10):925-926. [doi: 10.1136/jech-2014-204806] [Medline: 25480407]

Abbreviations

DPI: diabetes prevention intervention
GDM: gestational diabetes mellitus
MI: motivational interviewing
MRC: Medical Research Council
OGTT: oral glucose tolerance test
PE: process evaluation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus

Edited by T de Azevedo Cardoso; submitted 24.08.23; peer-reviewed by YL Leung; comments to author 26.01.24; revised version received 06.03.24; accepted 21.06.24; published 06.02.25.

Please cite as:
Sa'id IIM, Hotung N, Benton M, Nadal IP, Baharom A, Prina M, Nisak Mohd Yusof B, Goldsmith K, Birts S, Siew Mooi C, Forbes A, Ismail K, Chew BH
Process Evaluations of Interventions for the Prevention of Type 2 Diabetes in Women With Gestational Diabetes Mellitus: Systematic Review
Interact J Med Res 2025;14:e51718
URL: https://www.i-jmr.org/2025/1/e51718
doi:10.2196/51718
PMID:

Original Paper

Evolution of Artificial Intelligence in Medical Education From 2000 to 2024: Bibliometric Analysis

Rui Li[1], BSN; Tong Wu[2], MD, PhD

[1]Emergency Department, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Tong Wu, MD, PhD
Department of Obstetrics and Gynecology
Tongji Hospital, Tongji Medical College
Huazhong University of Science and Technology
1095 Jiefang Road
Wuhan, 430030
China
Phone: 86 13018044651
Email: tongwu66@tjh.tjmu.edu.cn

Abstract

Background: Incorporating artificial intelligence (AI) into medical education has gained significant attention for its potential to enhance teaching and learning outcomes. However, it lacks a comprehensive study depicting the academic performance and status of AI in the medical education domain.

Objective: This study aims to analyze the social patterns, productive contributors, knowledge structure, and clusters since the 21st century.

Methods: Documents were retrieved from the Web of Science Core Collection database from 2000 to 2024. VOSviewer, Incites, and Citespace were used to analyze the bibliometric metrics, which were categorized by country, institution, authors, journals, and keywords. The variables analyzed encompassed counts, citations, H-index, impact factor, and collaboration metrics.

Results: Altogether, 7534 publications were initially retrieved and 2775 were included for analysis. The annual count and citation of papers exhibited exponential trends since 2018. The United States emerged as the lead contributor due to its high productivity and recognition levels. Stanford University, Johns Hopkins University, National University of Singapore, Mayo Clinic, University of Arizona, and University of Toronto were representative institutions in their respective fields. *Cureus*, *JMIR Medical Education*, *Medical Teacher*, and *BMC Medical Education* ranked as the top four most productive journals. The resulting heat map highlighted several high-frequency keywords, including performance, education, AI, and model. The citation burst time of terms revealed that AI technologies shifted from imaging processing (2000), augmented reality (2013), and virtual reality (2016) to decision-making (2020) and model (2021). Keywords such as mortality and robotic surgery persisted into 2023, suggesting the ongoing recognition and interest in these areas.

Conclusions: This study provides valuable insights and guidance for researchers who are interested in educational technology, as well as recommendations for pioneering institutions and journal submissions. Along with the rapid growth of AI, medical education is expected to gain much more benefits.

(Interact J Med Res 2025;14:e63775) doi:10.2196/63775

KEYWORDS

artificial intelligence; medical education; bibliometric; citation trends; academic pattern; VOSviewer; Citespace; AI

Introduction

Artificial intelligence (AI) seeks to develop innovative machines that are capable of responding in ways similar to human intelligence. Its scope encompasses robotics, language recognition, image recognition, natural language processing, and expert systems. In recent years, the integration of AI into medical education (AIME) has significantly transformed traditional teaching and learning methodologies [1,2]. It can enhance clinical reasoning training, facilitate adaptive learning,

construct innovative medical teaching platforms, and facilitate critical care simulation teaching. For instance, the AI-powered simulation mannequins called SimMan 3G PLUS allows medical students to practice their clinical skills and teamwork in a fully controlled and immersive environment. Augmented reality and virtual reality technologies enable students to gain valuable hands-on experience by recreating real-life situations, without compromising patient safety [3]. In other cases, AI helps tailor educational content based on individual students' needs and abilities, and it not only fosters academic success but also cultivates critical thinking [1,4]. AI applications have become an important tool and situation for both educators and students.

As AI continues to evolve and advance, its role in promoting medical education is expected to grow. Despite many publications that have explored AI's application in medical research, they are limited to specific clinical questions in cardiology, dentistry, oncology, etc [5,6]. No bibliometric analysis has ever been conducted to assess medical education, which is essential for medical talent reserve. Therefore, a comprehensive analysis of current scientific literature is imperative for the continued advancement of AIME.

Bibliometrics uses mathematical and statistical methodologies to evaluate scholarly productivity and publication patterns within a specific discipline. We used bibliometrics to analyze the contemporary academic landscape, research priorities, and emerging trends in AIME in the 21st century. We revealed the knowledge clusters, scientific social patterns, and evolutionary nuances from an objective standpoint. This work not only simplifies navigation for researchers across disciplines but also gives valuable guidance to newcomers in the field, directing them toward promising avenues for future research.

Methods

Study Design

The research adhered to the step-by-step guidelines of bibliometric analysis and followed the reporting guideline of bibliometric reviews in biomedical literature [7].

Data Collection

The Web of Science Core Collection database was selected as the primary tool to identify relevant publications for this study. This database is well-known for its exceptional bibliometric research capabilities across over 190 subject areas and offers manual-checked literature retrieval. A 2-step process was used to construct an effective retrieval strategy. First, we extracted search terms from relevant systematic reviews and meta-analyses. Second, these search terms were meticulously reviewed and refined by authors (TW and RL) to ensure adequate coverage of all relevant research topics. The final search strategy was formulated based on this collaborative assessment: TS=("artificial intelligence" or AI or "convolutional neural network" or "recurrent neural network" or "long short-term neural memory" or "machine learning" or "genetic algorithm" or "evolutionary algorithm" or "artificial neural network" or "support vector machine" or "fuzzy logic" or "random forest" or "deep learning" or "natural language processing" or "speech recognition" or "computer vision" or "smart robot" or "video recognition" or "image recognition") and TS=(medical education or medicine education). Data collection took place on October 1, 2024, with publication years restricted from 2000 to 2024. No language restrictions were applied during the search process. Full records of retrieved papers were exported in plain text format.

Selection of Eligible Studies

Since search engines may yield results that do not fully align with the intended criteria, it is necessary to perform a manual screening process to identify the final included literature. To address this, two authors (TW and RL) independently screened the titles and abstracts of available studies based on the following exclusion criteria: (1) publications unrelated to AI. For example, some topics are associated with artificial insemination. (2) Publications unrelated to medical education, such as engineering curriculum and patient education. Notably, a majority of these studies have applied AI technologies to clinical questions and research, and only a small number pertained to medical education. Any discrepancies between the studies selected by the authors were resolved through a consensus meeting to reach a binding decision.

Data Cleaning and Preprocessing

The records from the Web of Science Core Collection database undergo rigorous selection criteria adhere to meticulous curation standards, and it partly enhances our research quality [8]. To further ensure accuracy and consistency, we performed a thorough data cleaning process. This involved the elimination of duplicate papers, which were identified through digital objective identifiers and study titles. The biblioshiny package in RStudio was used to standardize the names of authors and institutions. Any variations in author names were consolidated. As an author usually affiliates with institutions at different times, only the institution of the author at the time of publication is retained. Synonyms, aliases, and singular or plural forms, such as "AR," "augmented reality," and "augmented realities" were cleaned using a thesaurus file. The data cleaning was conducted manually by two authors (TW and RL).

Data Analysis and Visualization

To present the knowledge structure and emerging research trends, VOSviewer (version 1.6.18; Leiden University), Incites (Clarivate), InteractiVenn (Universidade de São Paulo), and Citespace (Drexel University) were used. Incites is a research evaluation tool developed by Clarivate Analytics, and the scores of citation impact, H-index, journal impact factor, international collaborations, study influence, and immediacy index were directly obtained in Incites. The citation impact is calculated by dividing the citation impact by the number of years since publication. The H-index indicates that H papers published by an author have been cited at least H times, thereby serving as a gauge of both scholarly productivity and influence. The journal impact factor was used based on the 2023 Journal Citation Reports. International collaborations were assessed to reveal the extent of interdisciplinary cooperation among coauthors hailing from various nations, highlighting the global reach and collaborative nature of the research topic. The study's influence measures the mean influence of a study within the first 5 years

of post publication, providing a view of its long-term impact. Concurrently, the immediacy index represents the average frequency with which a study is cited in the year of its publication, offering insight into the immediate reception in the academic community. The VOSviewer software was used to cluster countries, institutions, journals, and keywords. The Citespace software was used to identify keyword clusters and citation burst time. The specific parameter settings for analysis in VOSviewer and Citespace are provided in Tables 1 and 2. The latent semantic indexing and log-likelihood ratio algorithms were used for literature clustering. The InteractiVenn tool was used to identify the specific and overlapping journals in 4 categories: count, citation, top 10% papers, and cooperative work. To compare the frequencies of keywords in the AI and medical education fields, the keywords were classified based on their basic definition, and lists of professional terms or vocabulary or terminology.

Table 1. Information for clustering using the VOSviewer.

Item	Type	Algorithm	Normalization method	Minimum document or occurrence[a]
Country	Coauthorship	Majorization	Full counting	12
Organization	Coauthorship	Majorization	Full counting	15
Journal	Citation	Majorization	—[b]	12
Keyword	Co-occurrence	Majorization	Full counting	33

[a]We set no limit on the minimum citation.

[b]Not applicable.

Table 2. Information for keyword cluster using the Citespace.

Setting	Value
Time slicing	From January 2000 to October 2024, one year per slice
Text processing	Title, abstract, author keywords, and keywords plus
Node type	Keywords
Links	Strength: cosinel Scope: within slices
Selection criteria	k=7

Ethical Considerations

The institutional review board of Tongji Hospital deemed ethical approval unnecessary for this study.

Results

Baseline Characteristics

This study initially retrieved a total of 7534 publications from the search. Figure 1 provides a detailed flowchart illustrating the publication retrieval and selection process. After title and abstract screening, 2775 publications remained for further bibliometric analysis. The academic publications underwent a relatively flat growth in publication numbers from 2000 to 2017 (Figure 2A). However, the number of annual papers exhibited exponential growth since 2018, indicating the flourished development in this field. The citation counts also demonstrated a consistent rise, surpassing 10,000 citations in 2024 (Figure 2B). These documents encompassed original studies (n=1769, 63.75%), proceeding papers (n=467, 16.83%), reviews (n=237, 8.54%), and meeting abstracts (n=67, 2.41%). Proceeding papers typically contain the latest research findings and methods, as well as techniques within a scientific field, and their high percentage indicates that AIME is receiving much attention and developing rapidly. The research areas are primarily focused on Education Scientific Disciplines (n=415), Medicine General Internal (n=334), Health Care Sciences Services (n=323), Engineering Biomedical (n=264), and Surgery (n=255). It suggested that AIME is an interdisciplinary topic that needs support from specialized educators, clinical doctors, hospital administrators, and engineers.

Figure 1. Flowchart and basic information of retrieval strategy, screen process, and analysis methods.

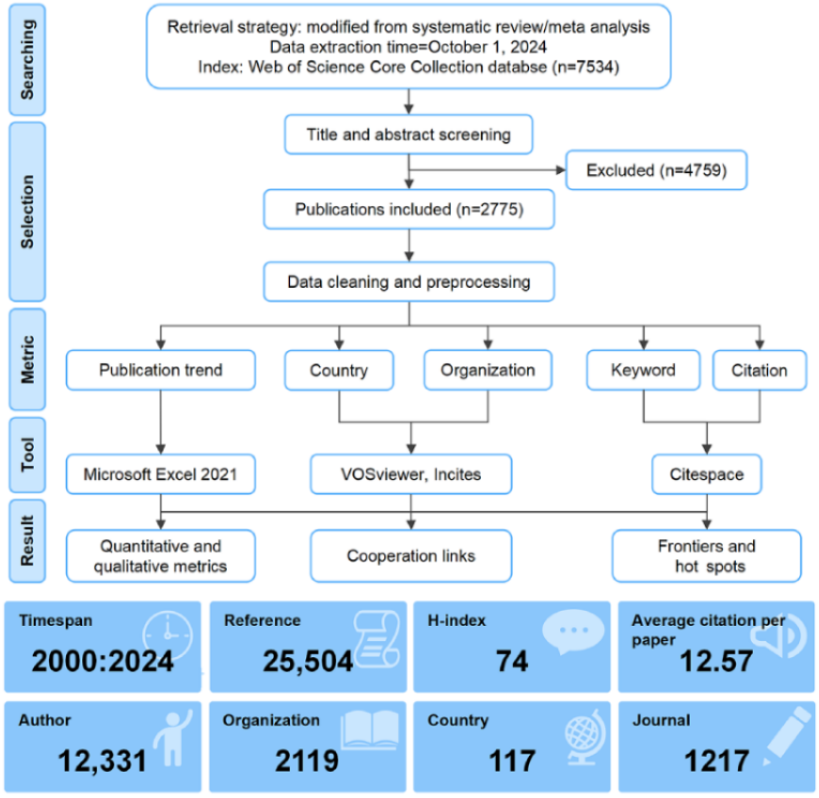
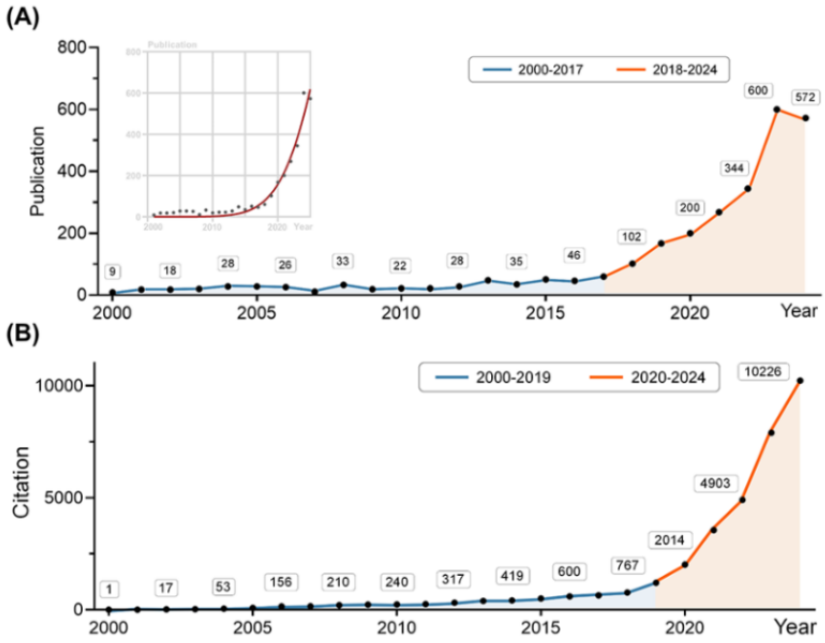


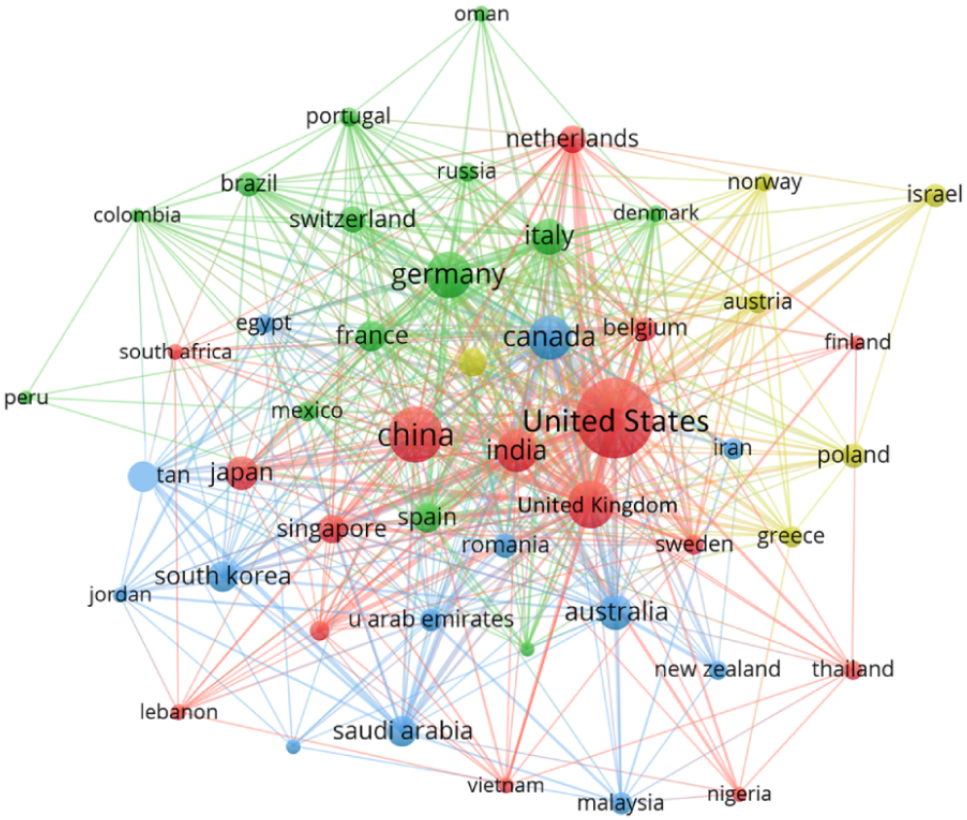
Figure 2. Baseline bibliometric characteristics of all eligible documents. (A,B) Line chart showing the annual number of documents and citations.



Productive and Influential Countries

In terms of national performance, 117 countries participated in the global discourse. The majority of research publications originated from Europe, North America, and Asia. The United States emerged as the primary contributor due to its high productivity (n=851 publications) and recognition levels (n=11,598 citations), exhibiting both the highest international and domestic collaborations. Although Chinese researchers published more papers than scholars from the United Kingdom (274 vs 227), their citations were relatively lower (3302 vs 3823). It may possibly be due to the limited innovation, restricted dissemination, and language barriers. The national cluster map revealed a diverse regional distribution pattern facilitated by these collaborative relationships (Figure 3). There was a strong collaboration among numerous countries, particularly close ties between the United States, China, the United Kingdom, and India.

Figure 3. Relationship and cluster of countries. Cooperative relationships networks among countries.



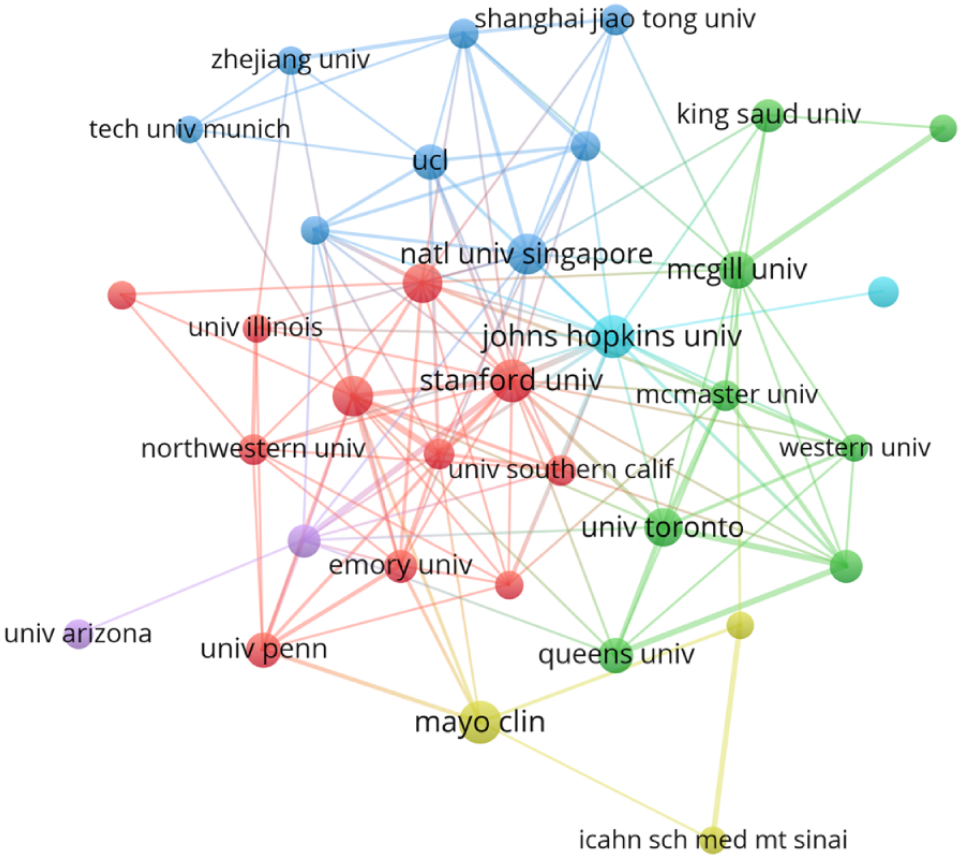
Institutional Performance

Altogether, 2119 institutions have participated in AIME studies. The top 10 institutions publishing the most papers are presented in Table 3. Among these, 4 institutions have published over 50 papers. Specifically, Harvard University emerged as the most prolific institution (n=70), and established the most international collaboration (n=35), while the University of London had the highest citation impact (n=27). The collaborative connections showed six clusters of countries publishing more than 40 papers (Figure 4). Stanford University, Johns Hopkins University, National University of Singapore, Mayo Clinic, University of Arizona, and University of Toronto were representative institutions in their respective cluster. Interestingly, regarding the impact relative to the world, the Alibaba Group had the highest score (value=61.43), followed by SUNY Downstate Health Sciences University (value=32.17), and Bukovinian State Medical University (value=25.14).

Table 3. Top 10 productive institutions of artificial intelligence in the medical education field.

Institution	Document	Citation	Citation Impact	International collaboration	H-Index	Impact relative to world
Harvard University	70	1085	15.50	35	17	1.25
University of California System	60	861	14.35	17	16	1.16
University of London	52	1211	23.29	27	17	1.88
Johns Hopkins University	51	1003	19.67	21	17	1.59
Stanford University	48	834	17.38	22	16	1.40
Harvard Medical School	47	588	12.51	25	13	1.01
Mayo Clinic	44	401	9.11	12	10	0.74
National University of Singapore	43	433	10.07	24	12	0.81
University of Toronto	39	624	16.00	14	14	1.29
University of Michigan	39	308	7.90	19	9	0.64

Figure 4. Visualization of institutional cooperation. Cooperative relationships networks.



Participants of the Source Journals

Cureus (n=78), *JMIR Medical Education* (n=72), *Medical Teacher* (n=54), and *BMC Medical Education* (n=54) ranked as the top 4 most productive journals among the 1217 involved journals (Figure 5A and Table 4). *JMIR Medical Education* published the most high-impact papers (n=48), and the second one, Cureus, only published 26. It is reasonable as *JMIR Medical Education* lies in the Q1 quartile, and likely receives more high-quality submissions, whereas Cureus is in the Q3 quartile. Cureus, *JMIR Medical Education*, *Medical Teacher*, and *International Journal of Computer Assisted Radiology and Surgery* were overlapped using a Venn diagram (Figure 5B). It was worth highlighting that Cureus and *JMIR Medical Education* had just begun to publish AIME-associated papers in 2020 (Figure 5C). Their open access characteristic helped them receive much attention in recent years. The publishers were diverse among the top 10 publication sources. More than half of these journals (7/10) published top 10% papers over 10. Notably, a significant proportion of the journals (n=1460, 52.61%) offered open access publishing, with gold access accounting for 33.26% (n=923) of this share.

Figure 5. Participated journals concerning artificial intelligence in the medical education field. (A) Relationship of representative journals. (B) Three significant journals were identified through a Venn diagram analysis of top publication volume, citation, top 10%, and cowork. (C) Annual academic output for the top five journals.

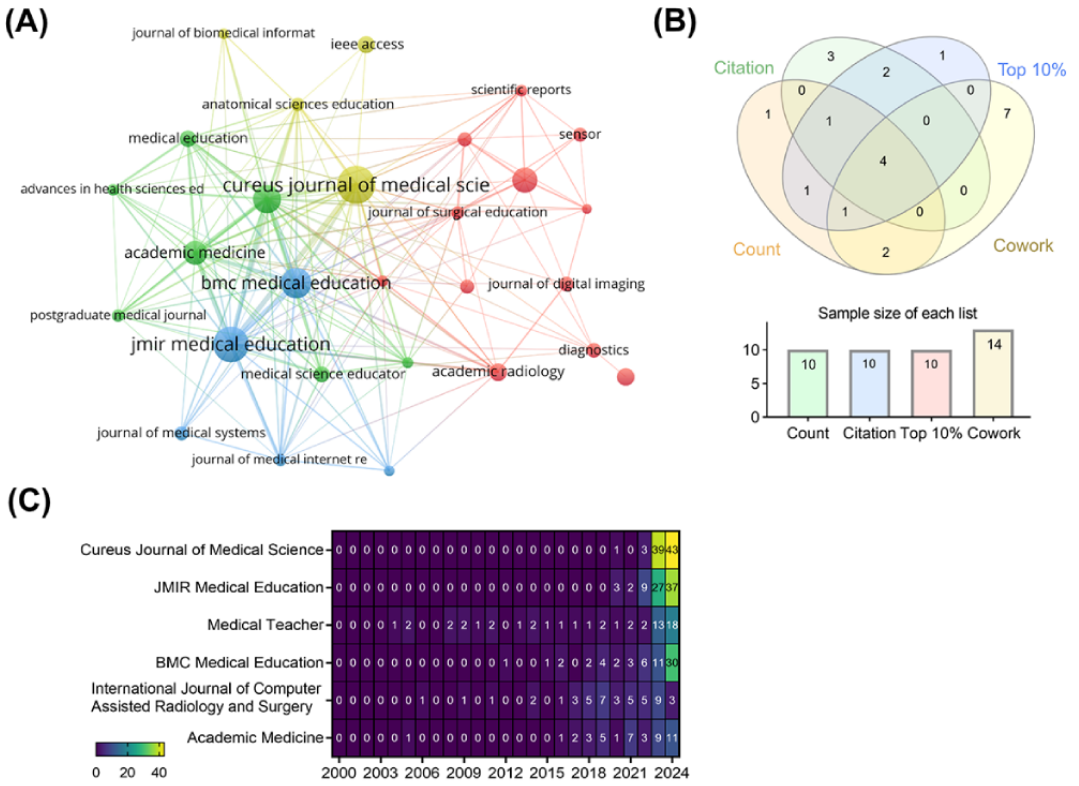


Table 4. Top five journals ranked by the counts of publications.

	Cureus	*JMIR Medical Education*	*Medical Teacher*	*BMC Medical Education*	*International Journal of Computer-Assisted Radiology And Surgery*
Document, n	78	72	54	54	46
Citation, n	910	751	1139	494	733
Top 10% paper	26	48	24	22	8
Publisher	Springer Nature	JMIR Publications Inc	Taylor & Francis Ltd	BMC	Springer Heidelberg
Cited half-life	2.5	2.1	9.1	4.4	4.8
Journal impact factor	1.1	3.2	4.7	3.4	3.1
Study influence	0.26	—[a]	1.53	0.86	0.74
Immediacy index	0.2	8.7	0.9	0.4	0.4
Open access	100%	97.22%	96.3%	24.07%	36.96%
JCI rank	173/329	18/85	10/85	15/175	67/204
JCI quartile	Q3	Q1	Q1	Q1	Q2

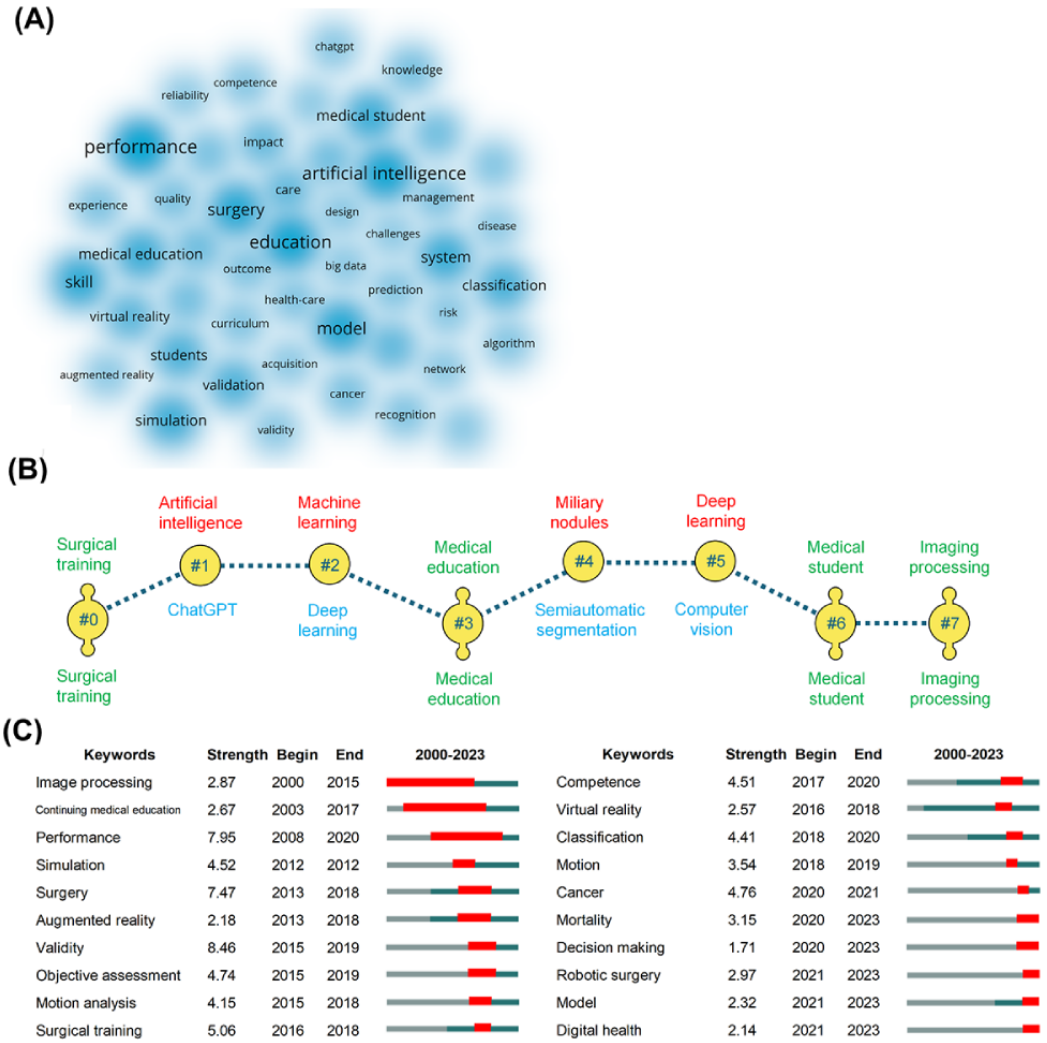
[a]Not applicable.

Keywords of Research Hot Spots

The study identified 7773 keywords from the titles and abstracts of the research materials, reflecting the central themes, areas of interest, and potential future developments within the discipline. The heat map highlighted high-frequency keywords, including performance (n=139), education (n=123), artificial intelligence (n=118), and model (n=103; Figure 6A). We next classified keywords according to the AI and medical education associated categories. The model (n=103), system (n=82), virtual reality (n=51), recognition (n=29), and algorithm (n=27) were popular topics, while most education focus was put on surgery (n=100), skill (n=90), simulation (n=75), classification (n=68), and diagnosis (n=47). The subsequent analysis produced eight distinct keyword clusters using log-likelihood ratio and latent semantic indexing algorithms (Figure 6B). Both algorithms

identified #0 surgical training, #3 medical education, #6 medical student, and #7 imaging processing as hub keywords. It suggested the critical role of AI applications in surgical operations and clinical image information. Citespace burst detection could reflect the research trends and innovative directions (Figure 6C). Notably, the citation burst time of terms revealed that AI technologies shifted from imaging processing (2000), augmented reality (2013), and virtual reality (2016) to decision-making (2020) and model (2021). Keywords such as mortality and robotic surgery persisted into 2023, suggesting the ongoing recognition and interest in these areas. Digital health involves the use of health apps, wearable equipment, and communication tools, which have been incorporated with AI. These devices not only provide personalized educational experiences and support mental health but also play a significant role in clinical teaching, disease prevention, and health promotion.

Figure 6. Keywords showing the hotspots and clusters. (A) Top keywords in the AIME field. (B) Eight clusters were obtained and unfolded using the latent semantic indexing (red), log-likelihood ratio (black) algorithms, or shared by both (green). (C) Keyword burstness panel showing the strength and duration. AIME: artificial intelligence into medical education.



Discussion

Principal Findings

AIME publication exhibits exponential growth these years, from approximately 50 counts in 2017 to 600 in 2023. This surge is further bolstered by the development of generative AI tools, and related national-level policies support. Here, we discuss the applications of AIME according to the keywords.

Machine Learning

Machine learning (ML), a subset of AI, is a technology that autonomously discerns inner patterns and relationships without explicit human instructions. For instance, after analyzing a large collection of cat and dog images, ML can identify distinguishing features, and subsequently differentiate between a cat and a dog in new photographs. The growing prominence of ML in medical education is evidenced by the surge in related topics, the approval of ML-based products, and the proliferation of entrepreneurial initiatives. ML-centric technologies have already been applied in clinical decision support systems, teaching tools, simulation, and training. An ML-based clinical decision support system could effectively integrate general information about HIV patients, including demographics, medical history, CD4+ lymphocyte count, viral load, genotypic data, and treatment history, to recommend an optimal combination antiretroviral

therapy [9]. More personalized advice on the appropriate dosage or duration of treatment will be provided with the help of other advanced ML algorithms [10]. In this context, it is vital to educate the next generation of medical professionals with adequate ML knowledge, enabling them to incorporate the outputs of ML tools into clinical decision-making, becoming part of this emerging data science revolution [11]. In the future, inexperienced medical students will use evidence-based learning models, like IBM's Watson Oncology system, as ordinary tools to interpret clinical data, make informed decisions, and recommend cancer treatments, in highly accordance with multidisciplinary teams [12].

Medical students require educational feedback to understand their performance and identify areas for improvement [13]. Traditional educational evaluation after surgical training falls short of providing timely, adequate, and objective assessment. A human rater is usually required to observe the video review to give a written or oral evaluation. The feedback is subjective and time-consuming, as well as limited to visual observation. With the help of automated ML-based AI assessments, the surgeon's performance can be objectively captured in a less resource-intensive way than human grading [14]. The motion, force, and vibration of robotic instruments are recorded according to the recognized structural metrics, like the Global Evaluative Assessment of Robotic Skill or the Objective Structured Assessment of Technical Skill [15].

Deep Learning

Deep learning (DL), a subset of ML, is often used to tackle complex tasks such as visual recognition, speech recognition, and natural language processing. This is achieved through the use of advanced architectures like convolutional neural networks, deep belief networks, and stacked auto-encoder networks. DL has been applied across medical undergraduate education, postgraduation education, and continuing education. A notable example is its use during medical retina rotations, where residents may not be fully trained due to limited time or access to complete their learning objectives. To address this, a DL-based model was developed to analyze a vast collection of retina images sourced from three public datasets, subsequently creating a comprehensive dataset for residents. These images were then distributed to each resident to aid in diagnosis, differential diagnosis, and therapeutic planning. The allocation system is tailored to each resident's case history, academic level, and performance, ensuring that those struggling with specific cases receive additional exposure to similar retinal conditions [16]. This approach enables AI models to identify the residents who will derive the most benefit from particular clinical cases, thereby significantly enhancing individualized ophthalmology education.

Additionally, DL technologies have been used to predict the difficulty of medical licensing examination questions, promoting more accurate assessments of examination difficulty. However, although DL models can effectively differentiate between cats and dogs, they do so by analyzing potentially hundreds or thousands of variables. The complexity and opacity of these variables often render them incomprehensible to humans. Consequently, there is an urgent need for improved interpretability methods in future DL applications to enhance understanding and transparency.

Natural Language Processing

Natural language processing plays a crucial role in smart health because of its ability to analyze and comprehend human language. The ChatGPT, developed by OpenAI Corporation, serves as a monumental tool for the natural language processing application. As for AIME, GPT-4 has been used for teaching cases, student analysis, creative writing, personalized learning guidance, and psychological support. ChatGPT is demonstrated to be effective in generating surgical procedure summaries, and its performance cannot be distinguished from a board-certified surgeon by less experienced residents [17]. In other cases, ChatGPT is used to teach the skills of breaking bad news [18], reasoning-based multiple-choice questions [19], and qualified examinations, including the Situational Judgement Test for final-year medical students in the United Kingdom [20]. These capabilities have revolutionized medical education and contributed to the overall improvement of health care delivery.

Segmentation

Image segmentation is a crucial AI technology in the radiology and pathology fields to accurately identify and delineate regions of interest, such as tumor lesions, ischemic tissues, and subcellular structures [21]. Traditional manual segmentation is not only time-consuming for physicians to learn and practice but also results in measurement variability that heavily depends on the observer's experience. In contrast, computer-assisted segmentation methods reduce the subjectivity and variability inherent in manual approaches, decrease processing time, and require minimal training. The use of image segmentation algorithms has led to significant improvements in the sensitivity and efficiency of detecting pulmonary nodules. Furthermore, these algorithms have been shown to enhance learning interest, bolster self-directed learning capabilities, sharpen problem-solving skills, and foster innovative thinking among medical trainees [22].

Effective surgical education for young surgeons presents significant challenges in practice. A segmentation-based system can identify key anatomical structures such as arteries, lymph nodes, and nerves during rectal cancer surgeries. Studies have shown the positive educational impact of AI-assisted videos in surgical training [23]. Some other AI technologies can facilitate surgical navigation and detect adverse events [24]. Taken together, real-time object segmentation is expected to play a major role in surgical education.

Comparison to Prior Work

AI has been increasingly used to enhance diagnostic accuracy, improve patient care, and facilitate the development of new treatments across various medical fields. In cardiology, for example, AI-powered algorithms are used to analyze electrocardiograms and detect arrhythmias, while in dentistry, AI-assisted diagnostic tools have been developed to identify oral pathologies [25,26]. Similarly, in oncology, AI-driven platforms are used to predict tumor growth and response to treatment [27]. Rather than concentrating on specific clinical questions, this study adopts a broader perspective to investigate

the role of AI in medical education as a whole. High-quality medical education is the key to improving the level of medical care. Through our bibliometric analysis, research trends and hot topics in this field were identified, offering insights into how AI is transforming medical education globally.

Future Direction

AIME has witnessed a remarkable evolution from the initial enthusiasm phase to the current acceptance situation [2]. Although the potential of AI technologies is expected to revolutionize medical education, related ethical considerations and challenges should be carefully examined. One major concern is whether AI might diminish the competence of medical students by increasing their reliance on external tools. Additionally, there is apprehension about the possibility of AI completely replacing medical educators. Issues such as bias, hallucinations, and uncertainties have further contributed to hesitancy in the acceptance and practical application of AI in the medical field [28]. It was necessary for a balanced approach to ensure sustainable implementation and find practical ways to incorporate AI into curricula. As deeper investigation is conducted, AI will be an integral part of medical education, highlighting a journey of personal and professional growth alongside technological adoption.

Since the release of ChatGPT, AI-generated content (AIGC) has emerged as an innovative educational tool with significant potential. AIGC technologies have the capability to reshape pedagogical practices through various applications, such as virtual patient construction, automated question bank generation, and 3D anatomical simulations [19,29]. These technologies may also help address the uneven distribution of educational resources and facilitate the updating of outdated knowledge. Exploring the application of AIGC in medical education represents both a transformation and expansion of existing teaching models, as well as a forward-looking exploration of future cultivation patterns.

AIME introduces new demands for educational accreditation systems and training models. While traditional medical education accreditation emphasizes faculty strength, physical facilities, and curriculum design, AI enriches these standards by incorporating digital classes, digital resources, and dynamic monitoring. In the context of postgraduate education, AIME necessitates a re-evaluation of training duration, content, and assessment methods. It is crucial for policy makers and practitioners to explore how AI can be reasonably integrated into residents' learning, ensuring a balance that prevents over-reliance on technology and maintains the importance of clinical practical experience.

Limitations

This study has several limitations. First, the bibliometric analysis was conducted exclusively using the Web of Science Core Collection database. This singular focus may introduce bias into the results, as it does not account for data from other significant sources. Future research should incorporate additional databases, such as Scopus and Google Scholar, to provide a more comprehensive analysis. Second, while citation frequency is often used as a measure of academic influence, it does not necessarily equate to positive evaluations. Citations can also be made for criticism or rebuttal. Therefore, supplementing bibliometric analysis with content analysis would offer a more accurate reflection of the literature's true impact. Finally, bibliometric analysis inherently concentrates on published research, which may not capture the most current research trends and developments.

Conclusions

This study delves into the current landscape of AI applications in medical education, encompassing the geographical distribution of research efforts, recognition of pivotal researchers, identification of key research trends, and exploration of emerging domains. There is a burgeoning interest in AIME and an expanding comprehension of its potential impact. The United States has emerged as a leader in this field, with many institutions standing out as prolific organizations. As the demand for more personalized and effective medical education grows, there is a pressing need for large-scale, rigorously designed studies to provide empirical evidence of AI's effectiveness and safety.

Acknowledgments

This work is financially supported by grants from the Teaching and Research Project of the Second Clinical College of Tongji Medical College (TJXYJ2023016 and TJSZ2024016). During the preparation of this work, the authors used Acadwrite [30] to improve readability, grammar, and language style. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

RL designed the study and wrote the initial draft of the manuscript. TW supervised the study, contributed to manuscript review and editing, and provided the funding. All authors contributed to the methodology and data visualization. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Cooper A, Rodman A. AI and medical education—a 21st-century pandora's box. N Engl J Med 2023;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
2. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME guide No. 84. Med Teach 2024;46(4):446-470 [FREE Full text] [doi: 10.1080/0142159X.2024.2314198] [Medline: 38423127]
3. Montoya-Rodríguez MM, de Souza Franco V, Llerena CT, Cobos FJM, Pizzarossa S, García AC, et al. Virtual reality and augmented reality as strategies for teaching social skills to individuals with intellectual disability: a systematic review. J Intellect Disabil 2023;27(4):1062-1084. [doi: 10.1177/17446295221089147] [Medline: 35420906]
4. Bhatnagar A, Kekatpure AL, Velagala VR, Kekatpure A. A review on the use of artificial intelligence in fracture detection. Cureus 2024;16(4):e58364 [FREE Full text] [doi: 10.7759/cureus.58364] [Medline: 38756254]
5. Lu W, Yu X, Li Y, Cao Y, Chen Y, Hua F. Artificial intelligence-related dental research: bibliometric and altmetric analysis. Int Dent J 2024 Sep 11 (forthcoming) [FREE Full text] [doi: 10.1016/j.identj.2024.08.004] [Medline: 39266401]
6. Suero-Abreu GA, Hamid A, Akbilgic O, Brown SA. Trends in cardiology and oncology artificial intelligence publications. Am Heart J Plus 2022;17:100162 [FREE Full text] [doi: 10.1016/j.ahjo.2022.100162] [Medline: 38559882]
7. Montazeri A, Mohammadi S, Hesari PM, Ghaemi M, Riazi H, Sheikhi-Mobarakeh Z. Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. Syst Rev 2023;12(1):239 [FREE Full text] [doi: 10.1186/s13643-023-02410-2] [Medline: 38102710]
8. Lim WM, Kumar S, Donthu N. How to combine and clean bibliometric data and use bibliometric tools synergistically: guidelines using metaverse research. J Bus Res 2024 Sep;182:114760. [doi: 10.1016/j.jbusres.2024.114760]
9. Parbhoo S, Bogojeska J, Zazzi M, Roth V, Doshi-Velez F. Combining kernel and model based learning for HIV therapy selection. AMIA Jt Summits Transl Sci Proc 2017;2017:239-248 [FREE Full text] [Medline: 28815137]
10. Peiffer-Smadja N, Rawson T, Ahmad R, Buchard A, Georgiou P, Lescure FX, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2020;26(5):584-595 [FREE Full text] [doi: 10.1016/j.cmi.2019.09.009] [Medline: 31539636]
11. James C, Wheelock KM, Woolliscroft JO. Machine learning: the next paradigm shift in medical education. Acad Med 2021;96(7):954-957. [doi: 10.1097/ACM.0000000000003943] [Medline: 33496428]
12. Jie Z, Zhiying Z, Li L. A meta-analysis of watson for oncology in clinical application. Sci Rep 2021;11(1):5792 [FREE Full text] [doi: 10.1038/s41598-021-84973-5] [Medline: 33707577]
13. González A. Turning a traditional teaching setting into a feedback-rich environment. Int J Educ Technol High Educ 2018 Aug 27;15(1):32. [doi: 10.1186/s41239-018-0114-1]
14. Brown JD, Kuchenbecker KJ. Effects of automated skill assessment on robotic surgery training. Int J Med Robot 2023 Apr;19(2):e2492. [doi: 10.1002/rcs.2492] [Medline: 36524325]
15. Nielsen PE, Foglia LM, Mandel LS, Chow GE. Objective structured assessment of technical skills for episiotomy repair. Am J Obstet Gynecol 2003;189(5):1257-1260. [doi: 10.1067/s0002-9378(03)00812-3] [Medline: 14634550]
16. Muntean GA, Groza A, Marginean A, Slavescu RR, Steiu MG, Muntean V, et al. Artificial intelligence for personalised ophthalmology residency training. J Clin Med 2023;12(5):1825 [FREE Full text] [doi: 10.3390/jcm12051825] [Medline: 36902612]
17. Lebhar MS, Velazquez A, Goza S, Hoppe IC. Dr. ChatGPT: utilizing artificial intelligence in surgical education. Cleft Palate Craniofac J 2024;61(12):2067-2073. [doi: 10.1177/10556656231193966] [Medline: 37545428]
18. Webb J. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. Cureus 2023;15(5):e38755 [FREE Full text] [doi: 10.7759/cureus.38755] [Medline: 37303324]
19. Agarwal M, Sharma P, Goswami A. Analysing the applicability of ChatGPT, Bard, and Bing to generate reasoning-based multiple-choice questions in medical physiology. Cureus 2023;15(6):e40977 [FREE Full text] [doi: 10.7759/cureus.40977] [Medline: 37519497]
20. Borchert RJ, Hickman CR, Pepys J, Sadler TJ. Performance of ChatGPT on the situational judgement test—a professional dilemmas-based examination for doctors in the United Kingdom. JMIR Med Educ 2023;9:e48978 [FREE Full text] [doi: 10.2196/48978] [Medline: 37548997]
21. Glotsos D, Spyridonos P, Cavouras D, Ravazoula P, Dadioti P, Nikiforidis G. Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models. Anal Quant Cytol Histol 2004;26(6):331-340. [Medline: 15678615]
22. Duong MT, Rauschecker AM, Rudie JD, Chen PH, Cook TS, Bryan RN, et al. Artificial intelligence for precision education in radiology. Br J Radiol 2019;92(1103):20190389 [FREE Full text] [doi: 10.1259/bjr.20190389] [Medline: 31322909]

23. Kinoshita K, Maruyama T, Kobayashi N, Imanishi S, Maruyama M, Ohira G, et al. An artificial intelligence-based nerve recognition model is useful as surgical support technology and as an educational tool in laparoscopic and robot-assisted rectal cancer surgery. Surg Endosc 2024;38(9):5394-5404. [doi: 10.1007/s00464-024-10939-z] [Medline: 39073558]
24. Bamba Y, Ogawa S, Itabashi M, Shindo H, Kameoka S, Okamoto T, et al. Object and anatomical feature recognition in surgical video images based on a convolutional neural network. Int J Comput Assist Radiol Surg 2021;16(11):2045-2054 [FREE Full text] [doi: 10.1007/s11548-021-02434-w] [Medline: 34169465]
25. Aldakhil S, Alkhurayji K, Albarrak S, Almihbash A, Aldalan R, Alshehri K, et al. Awareness and approaches regarding artificial intelligence in dentistry: a scoping review. Cureus 2024;16(1):e51825 [FREE Full text] [doi: 10.7759/cureus.51825] [Medline: 38327934]
26. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021;18(7):465-478 [FREE Full text] [doi: 10.1038/s41569-020-00503-2] [Medline: 33526938]
27. Hoang DT, Dinstag G, Shulman ED, Hermida LC, Ben-Zvi DS, Elis E, et al. A deep-learning framework to predict cancer treatment response from histopathology images through imputed transcriptomics. Nat Cancer 2024;5(9):1305-1317. [doi: 10.1038/s43018-024-00793-2] [Medline: 38961276]
28. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med 2018;378(11):981-983 [FREE Full text] [doi: 10.1056/NEJMp1714229] [Medline: 29539284]
29. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature 2023;614(7947):214-216. [doi: 10.1038/d41586-023-00340-6] [Medline: 36747115]
30. Acadwrite. URL: https://acadwrite.cn/ [accessed 2025-01-16]

Abbreviations

AI: artificial intelligence
AIGC: artificial intelligence–generated content
AIME: artificial intelligence into medical education
DL: deep learning
ML: machine learning

Edited by T de Azevedo Cardoso; submitted 29.06.24; peer-reviewed by W Yang, W Qi, Y Jiang; comments to author 19.09.24; revised version received 04.11.24; accepted 26.11.24; published 30.01.25.

Please cite as:
Li R, Wu T
Evolution of Artificial Intelligence in Medical Education From 2000 to 2024: Bibliometric Analysis
Interact J Med Res 2025;14:e63775
URL: https://www.i-jmr.org/2025/1/e63775
doi:10.2196/63775
PMID:

Original Paper

Increasing Participation and Completion Rates in Questionnaire Surveys of Primary Care Patients: Cluster-Randomized Study

Paul Sebo[1], MD, MSc; Benoit Tudrej[2], MD, PhD; Augustin Bernard[1], MD; Bruno Delaunay[1], MD; Alexandra Dupuy[1], MD; Claire Malavergne[1], MD; Hubert Maisonneuve[1,2], MD

[1]University Institute for Primary Care (IuMFE), University of Geneva, Geneva, Switzerland

[2]University College of General Medicine, Université Claude Bernard Lyon 1, Lyon, France

Corresponding Author:
Paul Sebo, MD, MSc
University Institute for Primary Care (IuMFE)
University of Geneva
Rue Michel-Servet 1
Geneva, 1211
Switzerland
Phone: 41 22 379 50 61
Email: paulsebo@hotmail.com

Abstract

Background: Participation and completion rates in questionnaire-based surveys are often low.

Objective: This study aims to assess participation and completion rates for a survey using paper and mixed mode questionnaires with patients recruited by research assistants in primary care waiting rooms.

Methods: This cluster-randomized study, conducted in 2023 in France, involved 974 patients from 39 practices randomized into 4 groups: “paper with incentive” (n=251), “paper without incentive” (n=368), “mixed mode with tablet” (n=187), and “mixed mode with QR code” (n=168). Analyses compared the combined paper group with the 2 mixed mode groups and the “paper with incentive” and “paper without incentive” groups. Logistic regressions were used to analyze participation and completion rates.

Results: Of the 974 patients recruited, 822 (women: 536/821, 65.3%; median age 52, IQR 37-68 years) agreed to participate (participation rate=84.4%), with no significant differences between groups. Overall, 806 patients (98.1%) answered all 48 questions. Completion rates were highest in the combined paper group (99.8%) compared to mixed mode groups (96.8% for paper or tablet, 93.3% for paper or QR code; P<.001). There was no significant difference in completion rates between the “paper with incentive” and “paper without incentive” groups (100% vs 99.7%).

Conclusions: Recruiting patients in waiting rooms with research assistants resulted in high participation and completion rates across all groups. Mixed mode options did not enhance participation or completion rates but may offer logistical advantages. Future research should explore incentives and mixed-mode strategies in diverse settings.

(Interact J Med Res 2025;14:e67981) doi:10.2196/67981

KEYWORDS

completion rate; missing data; mixed mode; web-based; participation rate; primary care; questionnaire; QR code; tablet; survey; primary care patients; randomized study

Introduction

Questionnaire-based surveys are valuable tools in a variety of research areas, including health care. For example, they allow data to be collected on patients’ experiences, opinions, and behaviors, which helps health care professionals to better understand their needs and improve the quality of care. Apart from the fact that they are generally inexpensive and take little time, 1 major advantage of questionnaire-based surveys lies in the flexibility they can offer in terms of the mode of administration (ie, face-to-face interviews, telephone interviews, and paper or web-based questionnaires completed at home or elsewhere).

However, to maintain a good level of representativeness and avoid selection bias, it is crucial to obtain high participation and completion rates. For surveys with postal, email, or telephone recruitment, several studies showed that participation rates (especially for web-based questionnaires) [1-4] and

completion rates (especially for paper questionnaires) [1,2] tended to be relatively low, whether the participants were individuals, patients, or physicians. Other studies showed that physicians' participation and completion rates were generally lower than those of the general population, probably mainly because of their workload in the practice and the increasing frequency with which they were asked to respond to surveys [5,6]. Although participation and completion rates are generally particularly low among physicians, these rates may also be suboptimal in studies with patients [7]. This means that the results of a large number of surveys conducted with patients may not be representative of the target population as a whole and may therefore lead to misleading conclusions.

Among the factors likely to influence participation and completion rates, the method chosen to collect data and the use of incentives are key elements [8-12]. Mixed modes, that is, the use of several methods in the same study, seem particularly useful for improving participation rates [9,13]. To our knowledge, QR codes have only been the subject of a few studies, focusing on specific populations and with limited sample sizes [11,14-16]. The usefulness of mixed modes including QR codes to improve participation and completion rates in a questionnaire has not yet been explored in general practice.

The aim of this study was to compare the participation and completion rates for a 48-question survey across 4 groups: a paper questionnaire group with incentive, a paper questionnaire group without incentive, and 2 mixed mode groups (ie, paper or web-based with tablet, and paper or web-based via QR code). Specifically, we analyzed differences between the combined paper group and the 2 mixed mode groups, as well as between the "paper with incentive" and "paper without incentive" groups. Patients were recruited into primary care waiting rooms by research assistants who were available to answer their questions if necessary. We hypothesized that both mixed mode options and incentives would improve participation and completion rates.

Methods

Design, Setting, and Study Population

This cluster-randomized study carried out in 2023 in the Rhône-Alpes region (France) was part of an environmental health project aimed at profiling different patterns of meat consumers in primary care, with a view to designing brief interventions to reduce meat consumption.

We used a professional register of primary care physicians from which we randomly extracted 200 physicians using computer-generated random numbers. Five research assistants contacted each randomly selected primary care physician by email until the required number of participating physicians (ie, n=40, Sample Size Determination section) was reached. If the primary care physician refused to participate or did not respond after 3 consecutive reminders, we contacted the next practice on the list.

The research assistants in this study were postgraduate residents in primary care medicine, conducting this work as part of their MD thesis. They had prior clinical experience, which enabled them to effectively interact with patients. To ensure consistency in patient recruitment and minimize variability, all research assistants underwent standardized training before the study began. This training included guidance on presenting the study, explaining the information sheet, and following ethical protocols. Although the study covered a large geographic area, requiring assistants to work autonomously during the recruitment phase, their approach was standardized and aligned with the study's protocols. Due to logistical constraints, the assistants were not randomly assigned to specific clinics.

We used simple randomization to allocate our sample of medical practices into 4 groups using computer-generated random numbers. Fifteen practices were randomized into the "paper without incentive" group, 10 into the "paper with incentive" group, 7 into the "paper or web-based with tablet" group (ie, patients were given the option of using either paper or a tablet provided by the researchers), and 7 into the "paper or web-based via QR code" group (ie, patients were given the option of using either paper or a QR code if they had their smartphones with them). We used computer-generated random numbers to determine the day of the week on which the study would be carried out in each practice.

Ethical Considerations

The study was approved by the Research Ethics Committee of the University College of General Practice, Claude Bernard Lyon 1 University (2023-01-03-01). The committee granted approval for the study protocol, including patient recruitment and data collection procedures. No additional ethical approvals were required beyond this institutional review. Participants were nonurgent, French-speaking, consecutive adult patients who were able to understand the study and provide written informed consent. They were informed about the study's objectives, data confidentiality, and their right to withdraw at any time without consequences. No personally identifiable information was collected, and all data were anonymized before analysis to ensure participant privacy and confidentiality. In line with the environmental health theme of the study, an origami paper containing a seed was provided as an unconditional incentive to all patients in the "paper with incentive" group, regardless of their willingness to complete the questionnaire. No financial compensation was offered to participants. No identifiable features of research participants are present in any images or supplementary materials included in the manuscript.

Data Collection

Participants were informed by a poster and recruited in the waiting room by a research assistant (20-25 patients per practice) between January 9, 2023, and June 16, 2023. The research assistant verified that the inclusion criteria were met and was available to answer any questions from the participants. The questionnaire (Multimedia Appendix 1) consisted of 48 questions: 4 sociodemographic questions (age, sex, postcode, and occupation), the French version of the feeling of the Sense of Coherence Scale (=17 questions), the French version of the Meat Attachment Questionnaire (=17 questions), and a questionnaire on intentionality adapted from previous studies (=10 questions) [17-19]. The questionnaire generally took about 10 minutes to complete. The paper and web-based questionnaires

were designed to be as similar as possible. In the 2 mixed modes (ie, with a tablet or QR code), we included an alert indicating that one or more questions were not answered. However, to be consistent with the paper version, participants did not have to answer all the questions before submitting the web-based questionnaire.

Statistical Analyses and Sample Size Determination

The sample size calculation was conducted to ensure sufficient power to detect differences in completion rates between groups. For simplification, the calculation was based on comparing proportions between 2 groups. Completion rates reported in the literature vary widely, ranging from 35% to 99%, depending on factors such as the study population, the theme of the study, and the format of the questionnaire [1,2,4]. We hypothesized that the completion rate in our study would be at least 80%, given the presence of a research assistant in the waiting room to assist participants if needed.

We calculated that 428 patients would be required in total to detect a 10-percentage-point difference between groups, with a power of 80% and a significance level of 5%, prior to accounting for clustering [20]. To adjust for clustering within practices, we assumed a conservative intraclass correlation coefficient of 0.02, based on findings from prior research in primary care settings [21], and an average cluster size of 20 patients per practice. This resulted in a design effect of 1.38, inflating the total adjusted sample size to 591 patients. With an average cluster size of 20 patients, this corresponded to 30 clusters in total. As 3 groups were compared, the sample size was increased to 40 clusters to ensure sufficient power.

We computed the proportion of patients who agreed to take part in the study (participation rate), the median number of questions answered with IQR, and the proportion of patients who answered all the questions (completion rate).

To assess whether the participation rate differed between groups, we used logistic regressions adjusted for intracluster correlation within practices. Comparisons were made between the combined paper group, the “paper or tablet” group, and the “paper or QR code” group as well as between the “paper with incentive” and “paper without incentive” groups. Multivariable analyses were not conducted for participation rates, as sociodemographic data were not collected for patients who refused to participate, in accordance with the ethics committee’s recommendations. This limitation prevented us from assessing variations in participation rates by sociodemographic factors.

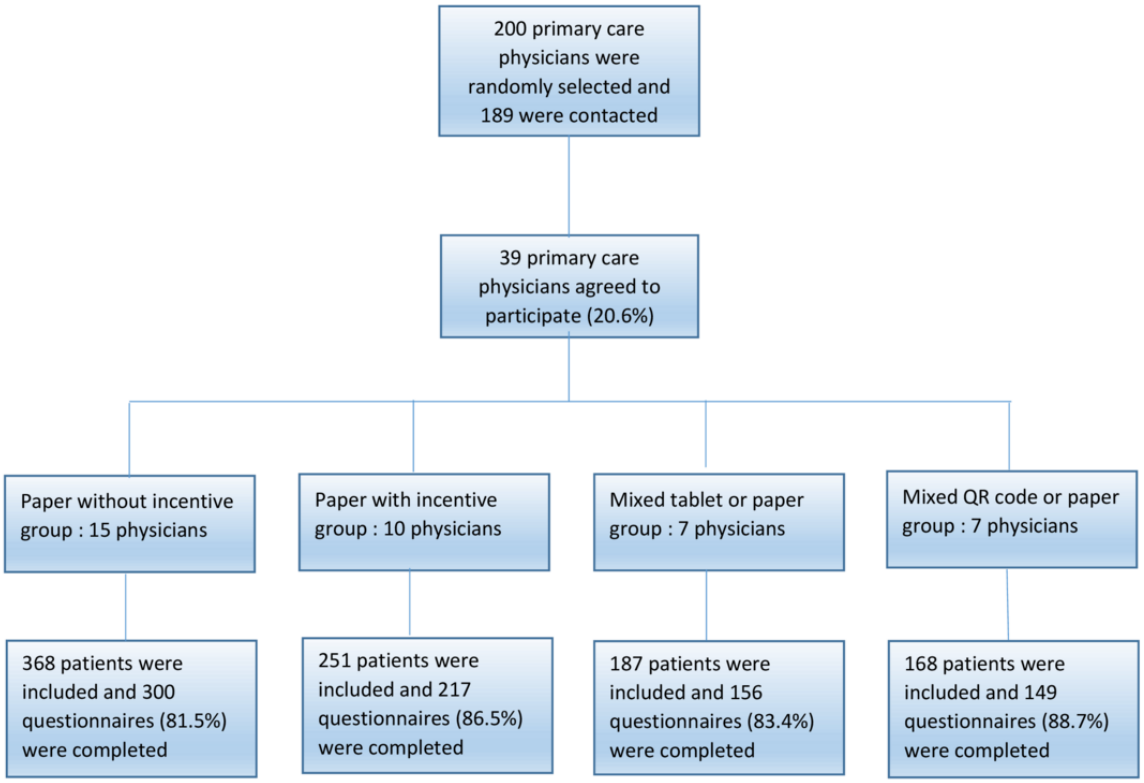
For the completion rate, exact logistic regressions adjusted for intracluster correlation were used to address sparse data issues, where cells formed by the outcome and categorical predictor variables had few or no observations [22]. Comparisons were made for (1) the three groups based on initial randomization (ie, combined paper, “paper or tablet,” and “paper or QR code”); (2) groups based on the questionnaire format chosen (ie, paper, “web-based with tablet,” and “web-based with QR code”); (3) the “paper with incentive” and “paper without incentive” groups; and (4) sex and age group (<40, 40-60, and ≥60). Multivariable analyses were also carried out, adjusting for sex, age group, and occupation.

All analyses were carried out with STATA (version 15.1; StataCorp LLC).

Results

The flowchart for the study is shown in Figure 1. From the list of 200 primary care physicians, 189 were contacted, of whom 39 (20.6%) agreed to participate. The study included 15 primary care physicians in the “paper without incentive” group, 10 in the “paper with incentive” group, 7 in the “paper or web-based with tablet” group, and 7 in the “paper or web-based via QR code” group.

Figure 1. Flowchart of the study.



The study sample consisted of 974 consecutive nonurgent patients who were recruited from these 39 medical practices, representing an average of 25 patients per practice (min=16, max=37). There were 251 patients in the "paper with incentive" group, 368 in the "paper without incentive" group, 187 in the "mixed mode with tablet" group, and 168 in the "mixed mode with QR code" group (Figure 1). Of these patients, 822 agreed to participate in the study, resulting in a participation rate of 84.4% (woman: 536/821, 65.3%; median age of 52, IQR 37-68; min-max=20-93; aged less than 40 years=239/819, 29.2%; 40-60 years=267/819, 32.6%; and more than 60 years=313/819, 38.2%). The distribution of women and men and by age group was similar in all 4 groups (*P* value=0.91 for sex and 0.08 for age group). The majority of participants were retirees (264/822; 32.1%), employees (227/822, 27.6%), and executives or intellectual professionals (166/822, 20.2%). Smaller groups included those not currently active in the workforce (68/822, 8.3%), intermediate professions (41/822, 5%), workers (28/822, 3.4%), artisans or business owners (25/822, 3%), and farmers (3/822; 0.4%).

The main results of the study are presented in Tables 1 and 2. Table 1 shows the differences between the combined paper group, the "mixed mode with tablet" group, and the "mixed mode with QR code" group. The data are organized first by the initial randomization group and then by the chosen questionnaire format. Table 2 focuses on comparing the "paper with incentive" group to the "paper without incentive" group.

Table 1. Participation and completion rates and median number of questions answered according to initial randomization group, chosen questionnaire format, sex, and age group[a].

	Participation rate, n/N (%)	Crude OR[b] (95% CI)	*P* value[c]	Number of patients having fully completed the questionnaire, n/N (%)	Median number of questions answered (IQR; min-max)	Crude OR (95% CI)	*P* value[d]	Adjusted OR (95% CI)	*P* value[e]
Initial randomization group (n=822)			.23				<.001		<.001
Paper (n=517)	517/619 (83.5)	Reference		516/517 (99.8)	48 (48-48; 47-48)	Reference		Reference	
Paper or tablet (n=156)	156/187 (83.4)	0.99 (0.55-1.80)		151/156 (96.8)	48 (48-48; 10-48)	0.05 (0-0.48)		0.06 (0-0.53)	
Paper or QR-code (n=149)	149/168 (88.7)	1.55 (0.91-2.63)		139/149 (93.3)	48 (48-48; 21-48)	0.02 (0-0.17)		0.03 (0-0.19)	
Chosen questionnaire format (n=822)							<.001		<.001
Paper (n=634)	N/A[f]	N/A	N/A	633/634 (99.8)	48 (48-48; 47-48)	Reference		Reference	
Web-based with tablet (n=105)	N/A	N/A	N/A	100/105 (95.2)	48 (48-48; 10-48)	0.03 (0-0.30)		0.03 (0-0.27)	
Web-based with QR code (n=83)	N/A	N/A	N/A	73/83 (88)	48 (48-48; 21-48)	0.01 (0-0.06)		0.01 (0-0.07)	
Sex (n=821)							.20		.17
Female (n=536)	N/A	N/A	N/A	529/536 (98.7)	48 (48-48; 10-48)	Reference		Reference	
Male (n=285)	N/A	N/A	N/A	277/285 (97.2)	48 (48-48; 21-48)	0.53 (0.17-1.62)		0.46 (0.14-1.47)	
Age group (years, n=819)							.06		.16
Less than 40 (n=239)	N/A	N/A	N/A	233/239 (97.5)	48 (48-48; 10-48)	Reference		Reference	
40-60 (n=267)	N/A	N/A	N/A	265/267 (99.3)	48 (48-48; 38-48)	5.06 (1.03-48.62)		3.99 (0.79-39.17)	
More than 60 (n=313)	N/A	N/A	N/A	308/313 (98.4)	48 (48-48; 21-48	2.35 (0.69-9.08)		1.92 (0.55-7.64)	

[a]The questionnaire consisted of 48 questions: 4 sociodemographic questions, the Sense of Coherence Scale (=17 questions), the Meat Attachment Questionnaire (=17 questions), and an intentionality questionnaire developed by our research team (=10 questions).

[b]OR: odds ratio.

[c]Univariable logistic regression (adjusted for intracluster correlation within practices).

[d]Univariable exact logistic regression (adjusted for intracluster correlation within practices).

[e]Multivariable exact logistic regression (adjusted for sex, age group, occupation, and intracluster correlation within practices).

[f]Participation rates are marked as "N/A" for the chosen questionnaire format, sex, and age group because participants who refused to participate did not indicate their preferred format, sex, or age. Therefore, participation rates, crude odds ratio, and *P* values by these characteristics could not be calculated.

Table 2. Comparison of participation and completion rates and median number of questions answered in the "paper with incentives" and "paper without incentives" groups[a].

	Participation rate, n/N (%)	Crude OR[b] (95% CI)	*P* value[c]	Number of patients having fully completed the questionnaire, n/N (%)	Median number of questions answered (IQR; min-max)	Crude OR (95% CI)	*P* value[d]
Paper group (n=517)	N/A[e]	N/A	.18	N/A	N/A	N/A	≥.99
Without incentives (n=300)	300/368 (81.5)	Reference		299/300 (99.7)	48 (48-48; 47-48)	Reference	
With incentives (n=217)	217/251 (86.5)	1.45 (0.84–2.49)		217/217 (100)	48 (48-48; 48-48)	1.74 (0-67.75)	

[a]The questionnaire consisted of 48 questions: 4 sociodemographic questions, the Sense of Coherence Scale (=17 questions), the Meat Attachment Questionnaire (=17 questions), and an intentionality questionnaire developed by our research team (=10 questions).

[b]OR: odds ratio.

[c]Univariable logistic regression (adjusted for intracluster correlation within practices).

[d]Univariable exact logistic regression (adjusted for intracluster correlation within practices).

[e]N/A: not applicable.

The participation rate ranged from 81.5% in the "paper without incentive" group to 88.7% in the "mixed mode with QR code" group, but the differences between the groups were not statistically significant. The incentive did not lead to a substantial increase in the participation rate. In this study, 51 patients (32.7%) in the "paper or web-based with tablet" group and 66 patients (44.3%) in the "paper or web-based via QR code" group preferred to complete the questionnaire using the paper version.

Overall, 806 patients (ie, 98.1% of patients) answered all 48 questions on the questionnaire. For the remaining 16 patients, the number of missing data ranged from 1 to 38. These patients were divided as follows: 1 patient belonged to the "paper without incentive" group, 5 to the "tablet" group, and 10 to the "QR code" group. The differences between groups based on the initial randomization and the questionnaire format chosen were small but statistically significant in both univariable and multivariable analyses. By contrast, the differences were not statistically significant for incentive, sex, and age group, but the number of patients with missing data was low.

Discussion

Main Findings

This study involved French primary care patients, recruited by research assistants and invited to complete a 48-question survey in the waiting room. We found that the participation rate was over 80% for all groups and that, when given the choice, a number of patients (n=117) preferred to complete the paper rather than the web-based questionnaire. We also found that the completion rate was 99.8% for the paper questionnaire and only slightly less for the web-based questionnaire. Finally, we found that the incentive had no influence on the participation rate or the number of questions answered.

Comparison With Existing Literature

Compared with other studies [12,23,24], we achieved excellent participation and completion rates in all groups (for participation rates, much higher than the 60% recommended in some publications [25]), despite a relatively lengthy self-administered questionnaire comprising 48 questions on a sensitive topic. The results of our study tend to reinforce the already known information that, among primary care patients, the response rate may not be affected by the length of the questionnaire, provided that the total duration is less than 15-20 minutes [26,27], which was the case in our study.

We found that the use of mixed options (paper or web-based), although intended to adapt more precisely to the preferences of each participant, did not appear to add value in terms of participation and completion rates. However, the web-based format (or mixed options) may be preferred for other reasons such as improved feasibility. Compared to paper questionnaires, the web-based format generally allows data to be obtained more cheaply (less printing, mailing, or typing costs), more quickly (real-time data tracking and less typing), and more accurately (the structured format minimizes incorrect entries and automatic data transfer minimizes data entry errors) [28]. Alongside the usual methods for collecting web-based data, QR codes have recently come into use in questionnaire surveys. They are simple and effective tools that are known to increase user engagement [29]. In a UK population-based maternity study, the inclusion of QR codes in the survey tended to lead to an increase in response rate, but this effect was limited and was probably also related to other factors (prior notification, short questionnaire, personalized study material, and additional reminder) [8]. In this English study, the initial response rate (30%) was considerably lower than ours.

The results of our study suggest that face-to-face recruitment with a research assistant achieves high participation and completion rates, providing support to participants. Previous studies have already highlighted that this strategy was associated with significantly higher participation rates [7,13,30]. Administering a face-to-face survey in the waiting room has the potential advantage of capturing patients' experiences when they are most accessible, with fewer competing demands than in other situations.

We found that, in the presence of a research assistant, neither the mixed mode, as discussed earlier, nor the unconditional nonmonetary incentive seemed to have any effect. Nonmonetary

incentives are recognized as important factors in improving the recruitment of participants to health-related surveys, with the impact being greater when the incentive is unconditional [31]. In our study, the high participation rate probably prevented us from observing a significant effect in the different methods tested. The presence of a research assistant seems to be the most powerful incentive, probably due to a desirability bias. This could be verified by carrying out the study using the incentives in the absence of a research assistant.

Significance and Recommendations for Future Research

This study makes a valuable contribution to the methodology of survey-based research in health care by underscoring the impact of recruitment strategies on both participation and completion rates. The notably high rates we observed across all groups suggest that face-to-face recruitment by a research assistant may be a critical factor for ensuring representativeness, especially in primary care settings where patient engagement can be challenging. Researchers conducting survey studies in similar environments may benefit from prioritizing in-person recruitment with trained personnel, as our findings indicate that this approach effectively mitigates common barriers to participation, such as lack of time. Additionally, while mixed mode survey options (including QR codes) did not significantly increase participation in this study, these tools still offer logistical advantages, such as ease of data collection, reduced error rates, and potential cost savings. For research teams looking to optimize both response rates and operational efficiency, mixed mode strategies may still be worthwhile. Future studies may benefit from investigating the impact of assistant-supported recruitment in different contexts or exploring alternative incentives, such as monetary incentives, that could further enhance response rates, particularly in larger or more geographically diverse samples.

Limitations

First, the results of our study cannot be generalized in all questionnaire-based surveys. Several studies showed that surveys carried out by recruiting participants by email, post, or telephone often led to mediocre participation rates for the web-based format [1-4] and mediocre completion rates for the paper format [1,2]. In this type of study, mixed mode options and incentives could be useful for positively influencing participation and completion rates. Second, selection bias is always possible in this type of survey, but in our opinion, it was reduced to a minimum by selecting physicians at random, including patients consecutively, and obtaining a high participation rate. In the context of our study, where patients were present in the waiting room, the selection biases usually described for electronic surveys in general practice do not apply [32]. Third, the study was carried out in only 1 region of France and only with primary care patients. The results obtained could have been different for other regions or countries, or for other study populations (physicians or patients visiting specialists). Finally, the use of simple randomization resulted in an unequal allocation of clusters across the study groups. While this does not affect the validity of randomization, it may have slightly reduced the power to detect differences between groups.

Conclusions

In conclusion, by recruiting patients in the waiting room with the help of a research assistant providing support to participants, we obtained a participation rate of over 80% in all groups. Neither the choice of response mode (ie, single or mixed, with or without a QR code) nor the use of incentives markedly influenced participation or completion rates. Future research would be useful to compare these rates with and without the presence of a research assistant in the waiting room to answer questions from participants. In terms of feasibility, the use of mixed options, including innovative methods such as QR codes, to offer participants the opportunity to choose, might be a relevant strategy, despite the results of our study showing no superiority in terms of participation and completion rates.

Acknowledgments

The authors of the study would like to sincerely thank all the primary care physicians and patients who collaborated on this project. The authors give special thanks to Amir Moussa for his involvement in this project. Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.

Data Availability

The data associated with this article are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire used in the study.
[DOCX File , 147 KB - ijmr_v14i1e67981_app1.docx]

References

1. Sebo P, Maisonneuve H, Cerutti B, Fournier JP, Senn N, Haller DM. Rates, delays, and completeness of general practitioners' responses to a postal versus web-based survey: a randomized trial. J Med Internet Res 2017;19(3):e83 [FREE Full text] [doi: 10.2196/jmir.6308] [Medline: 28330830]
2. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
3. Bergeson SC, Gray J, Ehrmantraut LA, Laibson T, Hays RD. Comparing web-based with mail survey administration of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group survey. Prim Health Care 2013;3(01) [FREE Full text] [doi: 10.4172/2167-1079.1000132] [Medline: 24078901]
4. McMahon SR, Iwamoto M, Massoudi MS, Yusuf HR, Stevenson JM, David F, et al. Comparison of e-mail, fax, and postal surveys of pediatricians. Pediatrics 2003;111:e299-e303. [doi: 10.1542/peds.111.4.e299] [Medline: 12671142]
5. Brtnikova M, Crane LA, Allison MA, Hurley LP, Beaty BL, Kempe A. A method for achieving high response rates in national surveys of U.S. primary care physicians. PLoS One 2018;13(8):e0202755 [FREE Full text] [doi: 10.1371/journal.pone.0202755] [Medline: 30138406]
6. Brehaut JC, Graham ID, Visentin L, Stiell IG. Print format and sender recognition were related to survey completion rate. J Clin Epidemiol 2006;59(6):635-641. [doi: 10.1016/j.jclinepi.2005.04.012] [Medline: 16713527]
7. Sitzia J, Wood N. Response rate in patient satisfaction research: an analysis of 210 published studies. Int J Qual Health Care 1998;10(4):311-317. [doi: 10.1093/intqhc/10.4.311] [Medline: 9835247]
8. Harrison S, Henderson J, Alderdice F, Quigley MA. Methods to increase response rates to a population-based maternity survey: a comparison of two pilot studies. BMC Med Res Methodol 2019;19(1):65 [FREE Full text] [doi: 10.1186/s12874-019-0702-3] [Medline: 30894130]
9. Booker QS, Austin JO, Balasubramanian BA. Survey strategies to increase participant response rates in primary care research studies. Fam Pract 2021;38(5):699-702. [doi: 10.1093/fampra/cmab070] [Medline: 34213547]
10. Parker MJ, Manan A, Urbanski S. Prospective evaluation of direct approach with a tablet device as a strategy to enhance survey study participant response rate. BMC Res Notes 2012;5(1):605 [FREE Full text] [doi: 10.1186/1756-0500-5-605] [Medline: 23110849]
11. Markwardt HS, Taghavi SE, Williams AP, Olivares MN, McDuffee PR, Hall BC. Tablet versus quick response code: Difference in research participation rates? J Technol Behav Sci 2023;8(2):178-181 [FREE Full text] [doi: 10.1007/s41347-023-00311-8] [Medline: 37215394]
12. Kelly BJ, Fraze TK, Hornik RC. Response rates to a mailed survey of a representative sample of cancer patients randomly drawn from the Pennsylvania Cancer Registry: a randomized trial of incentive and length effects. BMC Med Res Methodol 2010;10:65 [FREE Full text] [doi: 10.1186/1471-2288-10-65] [Medline: 20630101]
13. Anhang Price R, Quigley DD, Hargraves JL, Sorra J, Becerra-Ornelas AU, Hays RD, et al. A systematic review of strategies to enhance response rates and representativeness of patient experience surveys. Med Care 2022;60(12):910-918 [FREE Full text] [doi: 10.1097/MLR.0000000000001784] [Medline: 36260705]
14. Perez-Alba E, Nuzzolo-Shihadeh L, Espinosa-Mora JE, Camacho-Ortiz A. Use of self-administered surveys through QR code and same center telemedicine in a walk-in clinic in the era of COVID-19. J Am Med Inform Assoc 2020;27(6):985-986 [FREE Full text] [doi: 10.1093/jamia/ocaa054] [Medline: 32282922]
15. Turkistani KA. The use of innovative cost-saving audience response system in orthodontic case-based learning: a potential approach in distance learning. J Orthod Sci 2021;10:4 [FREE Full text] [doi: 10.4103/jos.JOS_44_20] [Medline: 34084760]
16. Masih EA. Nurse Educ Today 2022;111:105305 [FREE Full text] [doi: 10.1016/j.nedt.2022.105305] [Medline: 35220149]
17. Lacroix K, Gifford R. Reducing meat consumption: identifying group-specific inhibitors using latent profile analysis. Appetite 2019;138:233-241. [doi: 10.1016/j.appet.2019.04.002] [Medline: 30965045]
18. Prokop-Dorner A, Piłat-Kobla A, Zając J, Luśtyk M, Valli C, Łapczuk A, et al. Values and preferences influencing willingness to change red and processed meat consumption in response to evidence-based information: a mixed methods study. Public Health Nutr 2022;25(8):2084-2098. [doi: 10.1017/s1368980022000866]
19. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
20. Kirkwood BR, Sterne JAC, Kirkwood BR. Essential Medical Statistics. Hoboken, NJ: Blackwell Science; 2003.
21. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. J Clin Epidemiol 2004;57(8):785-794. [doi: 10.1016/j.jclinepi.2003.12.013] [Medline: 15485730]
22. Mehta CR, Patel NR. Exact logistic regression: Theory and examples. Stat Med 1995;14(19):2143-2160. [doi: 10.1002/sim.4780141908] [Medline: 8552893]
23. Kazzazi F, Haggie R, Forouhi P, Kazzazi N, Malata C. Utilizing the total design method in medicine: maximizing response rates in long, non-incentivized, personal questionnaire postal surveys. Patient Relat Outcome Meas 2018;Volume 9:169-172. [doi: 10.2147/prom.s156109]
24. Haller DM, Sanci LA, Patton GC, Sawyer SM. Text message communication in primary care research: a randomized controlled trial. Fam Pract 2009;26(4):325-330. [doi: 10.1093/fampra/cmp040] [Medline: 19542193]

25. Fincham JE. Response rates and responsiveness for surveys, standards, and the journal. Am J Pharm Educ 2008;72(2):43 [FREE Full text] [doi: 10.5688/aj720243] [Medline: 18483608]
26. Beebe TJ, Rey E, Ziegenfuss JY, Jenkins S, Lackore K, Talley NJ, et al. Shortening a survey and using alternative forms of prenotification: impact on response rate and quality. BMC Med Res Methodol 2010;10(1):50 [FREE Full text] [doi: 10.1186/1471-2288-10-50] [Medline: 20529365]
27. Guo Y, Kopec JA, Cibere J, Li LC, Goldsmith CH. Population survey features and response rates: a randomized experiment. Am J Public Health 2016;106(8):1422-1426. [doi: 10.2105/ajph.2016.303198]
28. Determann D, Lambooij MS, Steyerberg EW, de Bekker-Grob EW, de Wit GA. Impact of survey administration mode on the results of a health-related discrete choice experiment: online and paper comparison. Value Health 2017;20(7):953-960 [FREE Full text] [doi: 10.1016/j.jval.2017.02.007] [Medline: 28712625]
29. Cata T, Patel PS, Sakaguchi T. QR code: A new opportunity for effective mobile marketing. J Mobile Technol Knowl Soc 2013:ID748267 [FREE Full text] [doi: 10.5171/2013.748267]
30. Heerman WJ, Jackson N, Roumie CL, Harris PA, Rosenbloom ST, Pulley J, et al. Recruitment methods for survey research: Findings from the mid-south clinical data research network. Contemp Clin Trials 2017;62:50-55. [doi: 10.1016/j.cct.2017.08.006] [Medline: 28823925]
31. Zheng G, Oksuzyan S, Hsu S, Cloud J, Jewell MP, Shah N, et al. Self-reported interest to participate in a health survey if different amounts of cash or non-monetary incentive types were offered. J Urban Health 2018;95(6):837-849 [FREE Full text] [doi: 10.1007/s11524-018-0237-7] [Medline: 29654397]
32. Slater M, Kiran T. Measuring the patient experience in primary care: comparing e-mail and waiting room survey delivery in a family health team. Can Fam Physician 2016;62(12):e740-e748 [FREE Full text] [Medline: 27965350]

Edited by T de Azevedo Cardoso; submitted 25.10.24; peer-reviewed by A Dabbagh, M Sandri, SY Choi; comments to author 04.12.24; revised version received 26.12.24; accepted 17.01.25; published 25.02.25.

Please cite as:
Sebo P, Tudrej B, Bernard A, Delaunay B, Dupuy A, Malavergne C, Maisonneuve H
Increasing Participation and Completion Rates in Questionnaire Surveys of Primary Care Patients: Cluster-Randomized Study
Interact J Med Res 2025;14:e67981
URL: https://www.i-jmr.org/2025/1/e67981
doi:10.2196/67981
PMID:

Original Paper

Correlation Between Objective Habit Metrics and Objective Medication Adherence: Retrospective Study of 15,818 Participants From Clinical Studies

Antoine Pironet[1], PhD; L Alison Phillips[2], PhD; Bernard Vrijens[1], Prof Dr

[1]AARDEX Group, Seraing, Belgium

[2]Iowa State University, Ames, IA, United States

Corresponding Author:
Antoine Pironet, PhD
AARDEX Group
rue Bois St Jean 15/1
Seraing, 4102
Belgium
Phone: 32 43748640
Email: antoine.pironet@aardexgroup.com

Abstract

Background: Medication adherence, or how patients take their medication as prescribed, is suboptimal worldwide. Improving medication-taking habit might be an effective way to improve medication adherence. However, habit is difficult to quantify, and conventional habit metrics are self-reported, with recognized limitations. Recently, several objective habit metrics have been proposed, based on objective medication-taking data.

Objective: We aim to explore the correlation between objective habit metrics and objective medication adherence on a large dataset.

Methods: The Medication Event Monitoring System Adherence Knowledge Center, a database of anonymized electronic medication intake data from ambulant participants enrolled in past clinical studies, was used as the data source. Electronic medication intake data from participants following a once-daily regimen and monitored for 14 days or more were used. Further, two objective habit metrics were computed from each participant's medication intake history: (1) SD of the hour of intake, representing daily variability in the timing of medication intakes, and (2) weekly cross-correlation, representing weekly consistency in the timing of medication intakes. The implementation component of medication adherence was quantified using (1) the proportion of doses taken and (2) the proportion of correct days.

Results: A total of 15,818 participants met the criteria. These participants took part in 108 clinical studies mainly focused on treatments for hypertension (n=4737, 30%) and osteoporosis (n=3353, 21%). The SD of the hour of intake was significantly negatively correlated with the 2 objective adherence metrics: proportion of correct days (Spearman correlation coefficient, ρ_S=–0.62, P<.001) and proportion of doses taken (ρ_S=–0.09, P<.001). The weekly cross-correlation was significantly positively correlated with the 2 objective adherence metrics: proportion of correct days (ρ_S=0.55, P<.001) and proportion of doses taken (ρ_S=0.32, P<.001). A lower daily or weekly variability in the timing of medication intakes is thus associated with better medication adherence. However, no variability is not the norm, as only 3.6% of participants have 95% of their intakes in a 1-hour window. Among the numerous factors influencing medication adherence, habit strength is an important one as it explains over 30% of the variance in medication adherence.

Conclusions: Objective habit metrics are correlated to objective medication adherence. Such objective habit metrics can be used to monitor patients and identify those who may benefit from habit-building support.

(Interact J Med Res 2025;14:e63987) doi:10.2196/63987

KEYWORDS

medication adherence; compliance; habit; history; correlation; association; intake; electronic database; retrospective; medication; drug; adherence

Introduction

Medication adherence is "the process by which patients take their medication as prescribed" [1]. It comprises 3 components: initiation of the treatment, correct implementation of the prescribed regimen, and persistence to the treatment [1]. Poor medication adherence is a global public health issue [2,3] and has important negative consequences on the personal level [3-6], but also at the societal level [3-5]. Interventions aiming at improving medication adherence are abundant in the literature, but few have been shown to be effective across a population [5,6]. Most medication adherence interventions have focused on structural factors outside of the individual (simplification of the regimen, refill reminders, etc) or on behavior change interventions that target reflective or deliberative factors, such as patient education [7]. As at least half of our daily behaviors are nonreflective, but rather habitual [8], interventions that target these habitual processes may be more successful than education or persuasion-based interventions.

Habits are defined as automatic behaviors responding to recurring environmental cues [9]. As stronger medication-taking habit has been shown to be associated with better medication adherence [4,10-13], in particular its implementation component, some successful interventions have focused on improving habit [5,14]. However, medication-taking habit, and habit in general, is difficult to quantify [8]. Conventional habit metrics are self-reported [12], with an important example being the Self-Reported Habit Index [15], a 12-item questionnaire. Another example is the Self-Reported Behavioral Automaticity Index [16], a 4-item subset of the Self-Reported Habit Index. These and other self-reported indices endure common limitations associated with self-reported metrics, such as social desirability bias and poor patient recall [13].

Medication-taking habit can also be assessed using objective medication-taking data [17]. Such data is collected using smart medication packages, which can take several forms: an electronic cap fitted on a medication bottle [17], an inhaler with a chip embedded [13], a blister that detects when a pill is expressed out of a cavity, etc. The common feature of smart medication packages is that they passively timestamp each time a patient accesses their medication, thus providing objective data on when a patient takes their medication. This detailed information can be used to derive habit metrics quantifying the consistency of medication intake behavior over time [11,17]—the validity of which rests on the fact that habits are context-stable responses to conditioned cues. Such habit metrics are objective and do not endure the limitations associated with self-reported habit.

Day-to-day consistency of the timing of medication intake is frequently used as an objective habit metric. It has been operationalized as the variance of the hour of intake [17], its SD [13,18], or the proportion of medication intakes occurring in a fixed-size window, for instance, 2 [4,19,20], 3 [21,22], or 4 hours [23].

These day-to-day consistency measures will penalize someone for having different routines on different days of the week (for instance, systematically taking their medication at 6 PM on weekdays and 10 PM on weekends). However, this feature might not be desirable if the cue for medication-taking (eg, breakfast) is the same throughout the week, reflecting a good medication-taking habit. To overcome this limitation, Phillips et al [11] recently introduced another metric, called the weekly cross-correlation, based on the weekly consistency of intake timing. In this metric, the medication intake timing of each day of a given week is compared to the corresponding day of the next week.

Recently also, Hoo et al [13] introduced a "pragmatic habit index" empirically defined as the product of 2 variables: stability, measured as the SD of the hour of intake and frequency, measured as the proportion of prescribed doses taken. Introducing behavioral frequency in a habit index has been criticized [16], as it incorporates the dependent variable of interest (behavior frequency) in the predictors.

In a recent study on 79 patients with type 2 diabetes, objective habit metrics were found to correlate with objective medication adherence [11]. The goal of this paper is to reinvestigate this correlation on a 200-fold larger dataset covering multiple pathologies, introduced in the next section.

The main hypothesis is that a more consistent medication-taking habit is correlated to better medication adherence.

Methods

Data

This study was a retrospective cohort study. AARDEX Group's database, the Medication Event Monitoring System (MEMS; AARDEX Group) Adherence Knowledge Center was used as the data source. This database contains anonymized electronic medication adherence data from ambulant participants enrolled in clinical studies that ran between 1989 and 2016. These participants' medication adherence was electronically monitored using the MEMS. The following selection criteria were used: (1) once-daily medication with (2) a follow-up longer than 14 days. This second criterion was required because 14 days is the minimum duration needed to compute the weekly cross-correlation. These selection criteria matched data from 15,818 participants totaling 3,053,779 medication intakes.

Ethical Considerations

Human Participants' Ethics Review Approval or Exemptions

Approval from an institutional review board was not required for the present analysis, as it consists of secondary research for which the identity of the participants is unknown, per Title 45 of the Code of Federal Regulations Part 46, Subpart A, Section 46.104, Paragraph (d)(4)(ii) [24]. The original data collection for all studies included in this analysis was approved by institutional review boards.

Informed Consent

Data were obtained from participants enrolled in clinical studies. As such, the participants provided informed consent for the procedures of the original studies. These procedures required participants to store their medication in electronic medication

packages. Participants were informed that these electronic medication packages recorded their medication intakes. The original study procedures also included using the generated data to monitor the participants' dosing history and analyze their medication intake behavior. As a consequence, informed consent was not sought for this secondary analysis, because it corresponds to the use of the data that was originally presented to participants.

Privacy and Confidentiality

The data in AARDEX Group's database, the Adherence Knowledge Center, is anonymized.

Compensation Details:

Participants were compensated for their participation in the original clinical studies. Compensation modalities varied between studies and countries, and the authors do not possess information about the compensation process, which was managed by the sponsors of the original studies.

Objective Habit Metrics

For each participant, 2 objective habit metrics were computed. The first objective habit metric is the SD of the day-to-day hour of intake. The hours and minutes are extracted for each medication intake timestamp, irrespective of the date of intake, and the SD of the resulting list of hours and minutes is computed. The lower the SD, the more consistent a person is in the timing of their medication intakes. In the extreme case, when SD equals 0 h, all intakes occur at exactly the same time of the day.

The second objective habit metric is the "weekly cross-correlation" introduced by Phillips et al [11]. This metric compares each day of a given week with the corresponding day of the next week and quantifies whether intakes occurred around the same time on these 2 days. To do so, medication intakes are represented in a 2D matrix, denoted A, of dimensions $24 \times N_{days}$, with N_{days} being the number of follow-up days. The matrix is initially filled with zeros. Each medication intake occurring on day l at hour k is translated as a value of 1 for element $A[k,l]$. Then, each element $A[k,l]$ of the matrix is multiplied by its matching element from the previous week, $A[k,l-7]$. If intakes occur at the same time from week to week, the product will be close to 1. Otherwise, it will be close to zero. Finally, the weekly cross-correlation is equal to the sum of all element-wise products divided by the Euclidean norm of A. More details and code explaining how to compute these metrics are presented in Multimedia Appendix 1. As the pragmatic habit index incorporates adherence, it was not included in the present analysis.

Implementation Adherence Metrics

Medication adherence, more precisely its implementation component [1], was quantified as the proportion of correct days, that is, the proportion of days with exactly 1 intake. A second implementation metric was used: the proportion of doses taken. The proportion of doses taken was computed as the ratio between the total number of doses taken and the prescribed number of doses. In this work, since participants were on a once-daily regimen, the prescribed number of doses was equal to the number of follow-up days. If a participant discontinued treatment too early, the analysis was limited to the period during which the participant was on treatment.

The hypotheses were that (1) SD of the day-to-day hour of intake would be negatively correlated to objective adherence (the more variable the hour of intake, the worse the adherence); (2) the weekly cross-correlation would be positively correlated to objective adherence (the more consistent the pattern of intakes, the better the adherence); and (3) the weekly cross-correlation would be more strongly correlated to objective adherence than the SD of the hour of intake.

All dosing history data were compiled while patients were engaged with the monitored medication in the trial. Therefore, noninitiation and nonpersistence are not part of the adherence evaluation in this research.

Statistical Analyses

Continuous variables were reported using medians and first and third quartiles. Categorical variables were reported using counts and proportions. Pairwise correlations between continuous variables were assessed using the Spearman rank correlation coefficient. Correlation coefficients were converted using Fisher Z-transformation to obtain CIs [25,26]. Correlation coefficients were compared by performing a z-test on the difference between their Fisher Z-transformations [26,27]. The proportion of variance in adherence explained by habit was quantified using the Pearson correlation coefficient, squared [26]. Analyses were performed using Python 3 (Python Software Foundation) [28].

Results

The 15,818 participants whose data was extracted from the MEMS Adherence Knowledge Center took part in 108 clinical studies. These studies enrolled a median of 59 (IQR 22-169) participants. Table 1 presents the characteristics of the population.

Figure 1 presents the distributions and pairwise plots for the 4 variables of interest: proportion of correct days and proportion of doses taken as 2 objective measures of adherence, and SD of the hour of intake and weekly cross-correlation as 2 objective habit metrics. According to the 2 top-left diagonal panels, the proportion of correct days ranges between 0 and 1 by definition, while the proportion of doses taken ranges between 0 and 1.5 indicating that some participants took more doses than prescribed. The scatter plots also show that the proportion of doses taken is correlated to the proportion of correct days. In addition, the first is always superior to the second, which is a consequence of their definition.

According to the 2 bottom-right diagonal panels of Figure 1, the SD of the hour of intake ranges between 0 and 12 hours, while the weekly cross-correlation ranges between 0 and 1. These two findings are direct consequences of the definitions of these variables. According to the scatter plots, these 2 variables are negatively correlated, indicating that a larger consistency in day-to-day timing (low SD hour of intake) is associated with a larger consistency in week-to-week timing (high weekly cross-correlation).

Table 2 presents the associated correlation coefficients. All 4 Spearman correlation coefficients were significantly different from zero (T scores for the significance of the coefficients, from left to right and top to bottom: –61.48, –10.70, 104.11, and 48.59 with a df of 15,816, all P<.001).

According to Figure 1 and Table 2, the correlation between the SD of the hour of intake and objective adherence (measured using the proportion of correct days or proportion of doses taken) has a negative sign, indicating that, as hypothesized, the more day-to-day consistency in timing the higher the adherence. Figure 1 and Table 2 also show that the weekly cross-correlation is positively correlated to objective adherence, meaning that the more consistent a person's medication intake pattern is from week to week, the higher their medication adherence, which is also in line with the hypotheses.

The comparison of the Spearman correlation coefficients presented in Table 2 shows that the SD of the hour of intake is more strongly correlated to the proportion of correct days than the weekly cross-correlation is (95% CI for the difference between absolute values of Spearman coefficients: 0.05 to 0.09, z score for this difference=21.46, P<.001), conversely to what was hypothesized. On the other hand, the weekly cross-correlation is more strongly correlated to the proportion of doses taken than the SD hour of intake is (CI for the difference between absolute values of Spearman coefficients: 0.25 to 0.21, z score for this difference=9.45, P<.001), as hypothesized.

Habit accounts for about 30% of the variance in medication adherence, measured using the proportion of correct days: 29.18% for SD of the hour of intake (95% CI 29.17% to 29.19%) and 31.25% for the weekly cross-correlation (95% CI 31.24% to 31.26%). When medication adherence is measured using the proportion of doses taken, the proportion of variance explained by habit is smaller: 4.91% for the SD of the hour of intake (95% CI 4.91% to 4.92%) and 15.74% for the weekly cross-correlation (95% CI 15.73% to 15.75%).

Table 1. Population characteristics.

Pathology	Values
Hypertension, n (%)	4737 (30)
Osteoporosis, n (%)	3353 (21)
Viral hepatitis, n (%)	2005 (13)
Hypercholesterolemia, n (%)	1246 (8)
Angina, n (%)	968 (6)
AIDS, n (%)	754 (5)
Depression, n (%)	561 (4)
Diabetes, n (%)	549 (3)
Reversible airway obstruction, n (%)	397 (3)
Heart failure, n (%)	334 (2)
Attention-deficit/hyperactivity disorder, n (%)	301 (2)
Colorectal polyp, n (%)	192 (1)
Others, n (%)	421 (3)
Follow-up duration (days), median (IQR)	168 (60-365)
Proportion of correct days (%), median (IQR)	93.3 (83.7-97.7)
Proportion of doses taken (%), median (IQR)	100 (93.7-101.8)
SD hour of intake (h), median (IQR)	1.7 (1-2.7)
Weekly cross-correlation, median (IQR)	0.54 (0.38-0.69)

Figure 1. Pairwise relationships between the four variables of interest: 2 objective adherence metrics: proportion of correct days (number 1) and proportion of doses taken (number 2) and 2 objective habit metrics (numbers 3 and 4). Off-diagonal panels contain scatter plots for each pair of variables; on-diagonal panels show a histogram of the distribution of single variables.

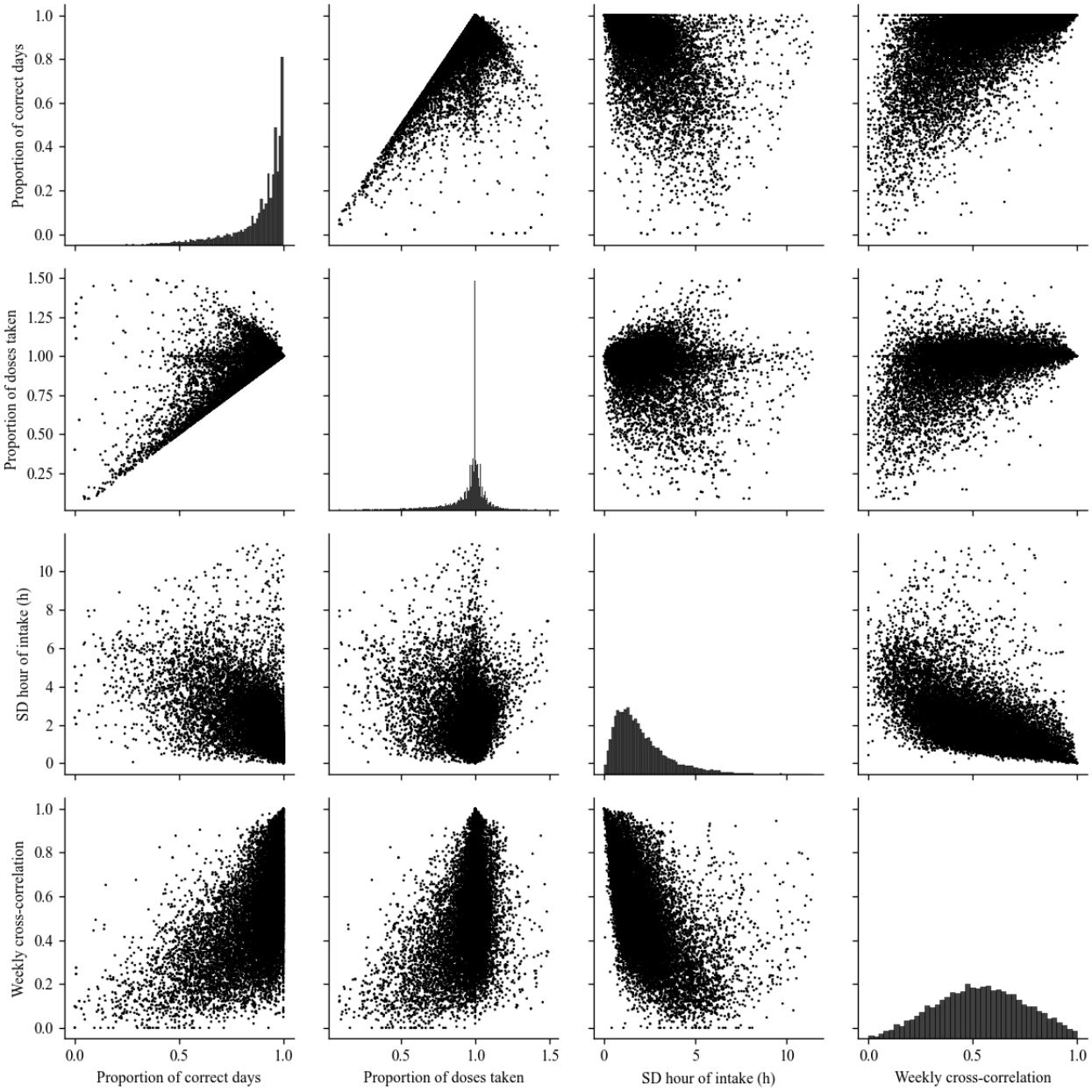


Table 2. Spearman correlation coefficients with 95% CI between 2 objective habit metrics (rows) and 2 objective adherence metrics (columns).

Objective habit	Objective adherence	
	Proportion of correct days	Proportion of doses taken
SD hour of intake	–0.62 (–0.63 to –0.61)	–0.09 (–0.1 to –0.07)
Weekly cross-correlation	0.55 (0.54 to 0.56)	0.32 (0.3 to 0.33)

Discussion

Principal Findings

This work showed that medication-taking habit is positively correlated to medication adherence in a very large, cross-pathology population of participants enrolled in clinical studies. Medication-taking habit was quantified using 2 measures of the consistency in the timing of medication intakes: day-to-day and week-to-week. The implementation component of medication adherence was quantified using 2 measures: the proportion of correct days and the proportion of doses taken. The conclusion was the same irrespective of the habit or adherence measure: the stronger the habit, the better the medication adherence.

The results showed no clear superiority of any of the 2 objective habit metrics per the strength of their correlation with adherence. The correlation of objective habit metrics with self-reported habit metrics, which are the most frequently used habit metrics, might be used to differentiate the 2 objective habit metrics. Previous studies exploring this correlation are discussed in the next section.

Comparison to Prior Work

The link between objective habit metrics and medication adherence has been investigated in very few other studies. The first one [11], performed by the authors on a much smaller dataset, reported very similar results: (1) a negative correlation between a measure of day-to-day variability and the proportion of correct days and (2) a positive correlation between the weekly cross-correlation and the proportion of correct days. The second one [13], previously discussed, used a habit metric incorporating adherence and reported an expected correlation between these 2 quantities.

Over 700 determinants of adherence have been reported [29]. Among all these determinants, objective habit strength seems to be an important one, as it explained over 30% of the variance in medication adherence in this work. This observation empirically justifies studying objective medication-taking habit.

A question related to the present work is whether people use time-based cues for taking their medications [10]. To answer this question, the proportion of participants having 95% of their intakes in a 1-hour window was computed. Only 3.6% of participants met this criterion. In a previous study [10], 21.7% of participants reported taking their medication at a specific time of the day. These numbers mean that a minority of patients rely on cues triggered by the clock and take medication at the same time every day. This finding might explain why reminders for medication adherence sent at a specific time of the day can be ineffective [30].

Few studies have studied how well objective habit metrics for medication adherence correlate with conventional self-reported habit metrics. SD or other metrics of day-to-day consistency were found to correlate with self-reported habit in 2 studies [17,18], but not in a third one [11]. Weekly cross-correlation was found to correlate with self-reported habits in 1 study [11].

Strengths and Limitations

First, the 2 objective habit metrics used in this study are measures of consistency in timing, day-to-day, or week-to-week. However, not all habits translate into consistency in timing, and consistency in timing can originate from other factors than habit [9]. In such cases, the objective habit metrics will not reliably quantify the presence or absence of habit. On the other hand, the objective consistency metrics have the advantage that they do not endure biases associated with self-reported habit metrics, such as recall bias.

Second, the dataset used in this paper was very large, so the findings are likely to be generalizable to the whole population of participants enrolled in clinical trials. On the other hand, no conclusion about a specific disease, population, or time can be drawn, because there would be a large confounding influence of the underlying study design and specific medication characteristics. The main result is that, in general, adherence is related to objective habit.

Third, in this study, habit and medication adherence were computed over a participant's whole follow-up, until treatment discontinuation. However, for a single participant, habit strength may have changed throughout the duration of this study, for instance, because of the introduction of a new device for medication adherence monitoring. The evidence on the effects that electronic adherence monitoring causes on adherence itself is unclear [22,31]. The effects of electronic adherence monitoring on medication-taking habit have never been studied, to the best of the authors' knowledge.

Another reason that may have caused individual changes in habit strength is that some of the studies analyzed were testing habit-building interventions. In such settings, the objective habit metrics used in this paper can be computed over shorter periods to capture the dynamics of habit formation and maintenance. For instance, in the case of a specific habit-building intervention, Pironet et al [32] observed that habit strength increased after a 6-month intervention and remained stable 6 months after the intervention was stopped.

Future Directions

If a patient's implementation adherence is measured electronically, their habit can be computed continuously from the recorded data and serve as a support for an intervention, in-person or through a mobile app. The intervention could focus on providing support to build a better medication-taking habit. An example of such a mobile intervention is presented in the study by Stawarz et al [33].

This work investigated the relationship between habit and the implementation component of medication adherence. Other, evenly important questions are whether objective habit indices predict the quality of implementation over the longer term, and if objective habit indices predict early treatment discontinuation.

Conclusion

In this study, the correlation between objective habit metrics, computed from electronic medication intake data, and objective medication adherence was assessed in a large database of participants enrolled in past clinical trials. The 2 objective habit metrics tested in this work were correlated to the 2 objective adherence metrics. These 4 pairwise correlations all imply that a lower variability in the pattern of medication intakes, be it day-to-day or week-to-week, is correlated with higher medication adherence. However, no variability is not the norm.

Objective habit metrics allow us to better identify patients who might benefit from habit-building support. If a patient uses electronic monitoring, the habit metrics can even be automated and computed in real time, directly reflecting habit changes and allowing for timely intervention. In addition, some effective interventions for medication adherence are based on developing or improving habits [5]. In such settings, electronic monitoring can be used to assess the effect of the habit-building intervention in real time [32].

Acknowledgments

The authors wish to thank Eric Tousset, MSc, from AARDEX Group for his help with the statistical analyses. No generative artificial intelligence model was involved in any portion of this work.

Data Availability

The individual values of objective medication adherence (proportion of correct days and proportion of doses taken) and objective habit (SD of the hour of intake and weekly cross-correlation) can be downloaded on AARDEX Group's website, in the section "Knowledge Center."

Authors' Contributions

AP handled the conceptualization, data curation, formal analysis, methodology, visualization, and writing of the original draft. BV worked on the conceptualization, methodology, supervision, visualization, and review and editing of the writing. LAP did the conceptualization, methodology, supervision, and review and editing of the writing.

Conflicts of Interest

AP is an employee of AARDEX Group. BV is a shareholder of AARDEX Group. AARDEX Group commercializes the MEMS, the hardware and software that was used to record the electronic medication adherence data used in this study.

Multimedia Appendix 1
Objective habit strength metrics.
[DOCX File , 18 KB - ijmr_v14i1e63987_app1.docx]

References

1. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, ABC Project Team. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
2. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency: adherence to medicines for diabetes, hypertension, and hyperlipidaemia. In: OECD Health Working Papers. Paris: OECD Publishing; 2018:105.
3. Jimmy B, Jose J. Patient medication adherence: measures in daily practice. Oman Med J 2011;26(3):155-159 [FREE Full text] [doi: 10.5001/omj.2011.38] [Medline: 22043406]
4. Phillips LA, Leventhal H, Leventhal EA. Assessing theoretical predictors of long-term medication adherence: patients' treatment-related beliefs, experiential feedback and habit development. Psychol Health 2013;28(10):1135-1151. [doi: 10.1080/08870446.2013.793798] [Medline: 23627524]
5. Russell CL, Hathaway D, Remy LM, Aholt D, Clark D, Miller C, et al. Improving medication adherence and outcomes in adult kidney transplant patients using a personal systems approach: SystemCHANGE™ results of the MAGIC randomized clinical trial. Am J Transplant 2020;20(1):125-136 [FREE Full text] [doi: 10.1111/ajt.15528] [Medline: 31291507]
6. Costa E, Giardini A, Savin M, Menditto E, Lehane E, Laosa O, et al. Interventional tools to improve medication adherence: review of literature. Patient Prefer Adherence 2015;9:1303-1314 [FREE Full text] [doi: 10.2147/PPA.S87551] [Medline: 26396502]
7. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
8. Wood W, Quinn JM, Kashy DA. Habits in everyday life: thought, emotion, and action. J Pers Soc Psychol 2002;83(6):1281-1297. [Medline: 12500811]
9. Volpp KG, Loewenstein G. What is a habit? Diverse mechanisms that can produce sustained behavior change. Organ Behav Hum Decis Process 2020;161:36-38. [doi: 10.1016/j.obhdp.2020.10.002]
10. Rajpura JR. Capsule commentary on Brooks et al., strategies used by older adults with asthma for adherence to inhaled corticosteroids. J Gen Intern Med 2014;29(11):1531 [FREE Full text] [doi: 10.1007/s11606-014-2962-2] [Medline: 25150031]
11. Phillips LA, Pironet A, Vrijens B. Evaluating objective metrics of habit strength for taking medications. J Behav Med 2023;46(4):632-641. [doi: 10.1007/s10865-023-00392-z] [Medline: 36662351]
12. Badawy SM, Shah R, Beg U, Heneghan MB. Habit strength, medication adherence, and habit-based mobile health interventions across chronic medical conditions: systematic review. J Med Internet Res 2020;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]

13. Hoo ZH, Wildman MJ, Campbell MJ, Walters SJ, Gardner B. A pragmatic behavior-based habit index for adherence to nebulized treatments among adults with cystic fibrosis. Patient Prefer Adherence 2019;13:283-294 [FREE Full text] [doi: 10.2147/PPA.S186417] [Medline: 30863018]
14. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: systematic review and meta-analysis. Prev Med 2017;99:269-276 [FREE Full text] [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
15. Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Soc Pyschol 2003;33(6):1313-1330 [FREE Full text] [doi: 10.1111/j.1559-1816.2003.tb01951.x]
16. Gardner B, Abraham C, Lally P, de Bruijn G. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
17. Phillips LA, Burns E, Leventhal H. Time-of-day differences in treatment-related habit strength and adherence. Ann Behav Med 2021;55(3):280-285 [FREE Full text] [doi: 10.1093/abm/kaaa042] [Medline: 32542355]
18. van de Vijver I, Brinkhof LP, de Wit S. Age differences in routine formation: the role of automatization, motivation, and executive functions. Front Psychol 2023;14:1140366 [FREE Full text] [doi: 10.3389/fpsyg.2023.1140366] [Medline: 37484115]
19. Linnemayr S, Odiit M, Mukasa B, Ghai I, Stecher C. Incentives and reminders to improve long-term medication adherence (INMIND): impact of a pilot randomized controlled trial in a large HIV clinic in Uganda. J Int AIDS Soc 2024;27(6):e26306 [FREE Full text] [doi: 10.1002/jia2.26306] [Medline: 38923298]
20. Phillips LA, Cohen J, Burns E, Abrams J, Renninger S. Self-management of chronic illness: the role of 'habit' versus reflective factors in exercise and medication adherence. J Behav Med 2016;39(6):1076-1091. [doi: 10.1007/s10865-016-9732-z] [Medline: 26980098]
21. Russell CL, Conn VS, Ashbaugh C, Madsen R, Hayes K, Ross G. Medication adherence patterns in adult renal transplant recipients. Res Nurs Health 2006;29(6):521-532 [FREE Full text] [doi: 10.1002/nur.20149] [Medline: 17131276]
22. McGrady ME, Ramsey RR. Using electronic monitoring devices to assess medication adherence: a research methods framework. J Gen Intern Med 2020;35(9):2707-2714 [FREE Full text] [doi: 10.1007/s11606-020-05905-z] [Medline: 32440997]
23. Stecher C, Mukasa B, Linnemayr S. Uncovering a behavioral strategy for establishing new habits: evidence from incentives for medication adherence in Uganda. J Health Econ 2021;77:102443 [FREE Full text] [doi: 10.1016/j.jhealeco.2021.102443] [Medline: 33831632]
24. Exemptions (2018 requirements). Office for Human Research Protections. 2019. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-a-46104/index.html [accessed 2025-01-24]
25. Altman DG. Practical Statistics for Medical Research. Boca Raton, FL: CRC Press; 1999:293-295.
26. Cohen J, Cohen P, West S, Aiken L. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2003:49-50.
27. Myers L, Sirois MJ. Spearman correlation coefficients, differences between. In: Encyclopedia of Statistical Sciences. Hoboken, NJ: Wiley-Interscience; 2006:7901-7903.
28. Van Rossum G, Drake FL. Introduction to Python 3: Python Documentation Manual Part 1. Scotts Valley, CA: CreateSpace; 2009.
29. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. Front Pharmacol 2013;4:91 [FREE Full text] [doi: 10.3389/fphar.2013.00091] [Medline: 23898295]
30. Liu X, Lewis JJ, Zhang H, Lu W, Zhang S, Zheng G, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med 2015;12(9):e1001876 [FREE Full text] [doi: 10.1371/journal.pmed.1001876] [Medline: 26372470]
31. Acosta FJ, Ramallo-Fariña Y, Bosch E, Mayans T, Rodríguez CJ, Caravaca A. Antipsychotic treatment dosing profile in patients with schizophrenia evaluated with electronic monitoring (MEMS®). Schizophr Res 2013;146(1-3):196-200. [doi: 10.1016/j.schres.2013.02.016] [Medline: 23474024]
32. Pironet A, Bartlett Ellis R, Stephen MB, Yerram P, Wakefield M, Awopetu D, et al. The SystemCHANGE intervention improves medication-taking habit. Germany: Springer; 2024 Presented at: Proceedings of the 27th annual meeting of ESPACOMP, the International Society for Medication Adherence; 2023 November 30-December 1; Budapest, Hungary p. 328-358. [doi: 10.1007/s11096-023-01688-5]
33. Stawarz K, Cox A, Blandford A. Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. : ACM; 2014 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2014 April 26-May 1; Toronto Ontario Canada p. 2269-2278. [doi: 10.1145/2556288.2557079]

Abbreviations

MEMS: Medication Event Monitoring System

Edited by T de Azevedo Cardoso; submitted 05.07.24; peer-reviewed by B Gardner, I Ghai, S Linnemayr; comments to author 12.09.24; revised version received 04.11.24; accepted 08.01.25; published 06.02.25.

Please cite as:
Pironet A, Phillips LA, Vrijens B
Correlation Between Objective Habit Metrics and Objective Medication Adherence: Retrospective Study of 15,818 Participants From Clinical Studies
Interact J Med Res 2025;14:e63987
URL: https://www.i-jmr.org/2025/1/e63987
doi:10.2196/63987
PMID:

Original Paper

Perceptions and Experiences of Caregiver-Employees, Employers, and Health Care Professionals With Caregiver-Friendly Workplace Policy in Hong Kong: Thematic Analysis

Maggie Man-Sin Lee[1], BBA, JD, MPH, PhD; Eng-kiong Yeoh[1], MBBS; Eliza Lai-Yi Wong[1], BSN, MPH, PhD, RN; Xue Bai[2], BL, BA, PhD; Nelson Chun-Yiu Yeung[1], BSSc, MPhil, PhD; Catherine French[3], BBA, MA, PhD; Henock Taddese[4], BSc, MEd, MPH, PhD

[1]The Jockey Club School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Department of Applied Social Science, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]Department of Health and Life Sciences, King's College London, London, United Kingdom

[4]School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom

Corresponding Author:
Henock Taddese, BSc, MEd, MPH, PhD
School of Public Health
Faculty of Medicine, Imperial College London
St Mary's Campus
167 Medical School
London, W2 1NY
United Kingdom
Phone: 44 02075949475
Email: h.taddese@imperial.ac.uk

Abstract

Background: Caregiver-employees (CEs) for older adults experience a high burden to fulfill their dual roles. Caregiver-friendly workplace policy (CFWP) has been used in many countries to balance employment and caregiving duties, but it is a relatively new concept in Hong Kong.

Objective: This study explored the views and experiences of CEs, employers, and health care professionals regarding CFWP (specifically for older adult caregivers) in Hong Kong.

Methods: This study explored the CFWP-related views and experiences in Hong Kong using 15 in-depth interviews with purposively sampled CEs for older adults, employers, and health care professionals.

Results: Two context-related themes ("lacking leadership" and "unfavorable culture") were identified with thematic analysis. They explain the absence of CFWP in Hong Kong due to the lack of governmental and organizational leadership, and the additional burden experienced by CEs because of the working culture that underpins work-life separation, overprizing business interest, and unsympathetic corporate attitude. Implicit voice theory was applicable in explaining CEs' nondisclosure about their status at work due to potential risks. In addition, the two facilitation-related themes ("role struggle" and "inadequate support") identified in this study exhibit how the dual role had positive and negative spillover effects on each other and the inadequacy of social welfare and health care support systems.

Conclusions: We strongly recommend exploring and adopting potential CFWP in Hong Kong, considering the complexity of factors identified in this study.

(Interact J Med Res 2025;14:e58528) doi:10.2196/58528

KEYWORDS

caregiver employees; workplace; discrimination; dual roles; caregiver burden

Introduction

Background

In 2020, one billion people globally were aged 60 years or older, which is expected to double by 2050 [1]. This unprecedented increase in the aging population poses numerous economic, political, social, and health care challenges [2]. One of the pressing issues has been providing daily care to older adults. However, the size of the professional and trained workforce has not increased proportionately to meet this demand [3], driving innovations in other fields, such as robotics, to relieve some of the unmet demands of older people care [4]. Even in countries where welfare systems are well established, family caregivers provide most of the care and support for older adults [5].

Consequently, caregivers experience high burdens, strain, and poor mental health outcomes such as burnout, anxiety, and depression, as noted by several studies [6,7]. Many caregivers may simultaneously engage in paid employment, referred to as caregiver-employees (CEs), who may experience additional burdens due to high professional and caregiving demands [8]. CEs experience significant impairments [7]; they are three times more vulnerable to adverse health issues than non-CEs [9]. The progressive decline in care recipients' functional capacity requires higher physical effort and caregiving time, resulting in poor health and depression among CEs [10].

Health care professionals (HPs) have traditionally provided psycho-educational support to promote caregivers' competency and well-being [11]. In addition, employers (ERs) are increasingly adopting caregiver-friendly workplace policies (CFWPs) to mitigate some of the caregivers' burden [7]. For instance, about 80% of ERs in the United States provide some CFWP [12]. CFWP typically includes flexible working arrangements, support services, and paid or unpaid leave [13] to help CEs manage their multiple roles and improve their work-life balance [7].

The growing international commitment to sustainable organizational behavior in which employee well-being is a significant determinant of their productivity, and therefore, organizational performance has accelerated the development and widespread adoption of CFWP [14-16]. For instance, studies have shown that in addition to employee well-being, CFWP may be critical for promoting the productivity of CEs [17] as they have been associated with improved overall health of CEs by reducing occupational and overall stress, minimizing work interruptions, and improving performance [18]. There are also direct economic benefits accruing from adopting CFWP. For example, educating CEs about their caregiving activities generates a net benefit ranging from US $48,010 to US $675,657 for CEs and ERs [19].

However, the three core stakeholders of CFWP may have different perspectives on the objectives, gaps, and limitations of the existing policies against the context of, for example, personal needs (for CEs), imparting caregiving competence (for HPs), or improving organizational performance (for ERs). Thus, exploring CFWP-related experiences and perceptions of CEs' and other stakeholders, especially HPs and ERs, is imperative. While several studies have reported the types and impact of CFWP [13,17], there is a paucity of studies reporting stakeholders' perspectives on CFWPs. Although we could not identify any study exploring the perspectives of CEs, one study explored the perspectives of managers working in the Canadian health care sector [20]. In addition, we could not identify any academic literature, public policy, or articulated organizational policy specifically devised for CEs caring for older adults.

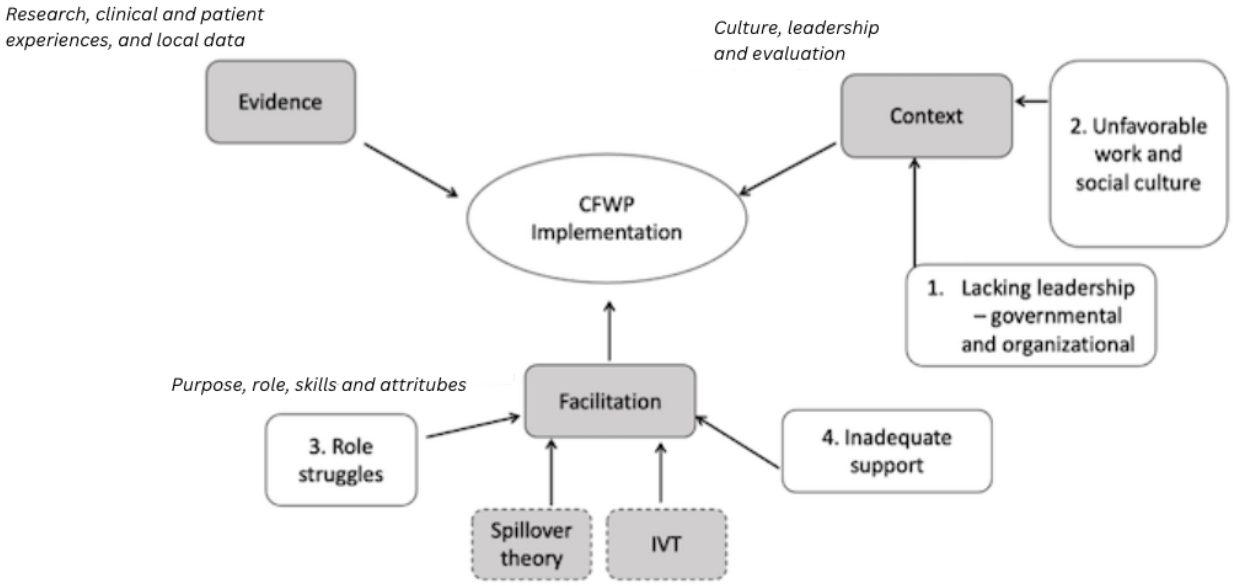
Therefore, this study explored the CFWP-related views and experiences of CEs from Hong Kong caring for older adults, along with the views and experiences of local HPs and ERs with prior experience with formulating or implementing CFWP.

Conceptual Framework

Since CFWP is a relatively new idea in Hong Kong, we adopted the Promoting Action on Research Implementation in Health Services (PARIHS) framework for this explorative qualitative study. PARIHS was developed by the Royal College of Nursing in the United Kingdom as a conceptual framework of interacting elements to implement evidence-based practices [21]. The framework's key constructs, namely evidence, context, and facilitation, were used to guide the interview with participants regarding any future policy implementation.

To further add to the framework and explicate the facilitating factors, spillover theory [22] and implicit voice theory (IVT) [23] were used. The two theories orientate the questions from the point of view of the CEs, whereby spillover theory explores whether the interface between the microsystems of work and family is positive or negative [24], and IVT takes the gaze toward any perceived risks of inappropriateness of speaking up or disclosure of status in an organizational hierarchy [25]. These theories are fundamental in exploring the facilitation of CFWP, in other words, the expectation of and potential actions of stakeholders vis-a-vis the relatively new policy idea. Spillover theory aids the exploration of the interactions of the dual roles—work and caring—while IVT sheds light on whether CEs could or would disclose their CE status and advance their interests. Subsequently, the guiding theoretical framework adapted for this study effectively superimposes IVT and spillover theories on top of the PARIHS framework (Figure 1).

Figure 1. Schematic diagram outlining the conceptual framework of the study. CFWP: caregiver-friendly workplace policy; IVT: implicit voice theories.



Methods

Study Setting

The economic and sociocultural factors in Hong Kong put carers in a particularly precarious position. Economically, laissez-faire has long been the foundation of Hong Kong's stability and prosperity [26,27]. However, free-market neoliberalism has failed Hong Kong in terms of increasing social inequality [28,29], leaving the economically and socially vulnerable even more susceptible to discrimination and unfair treatment. In our case, the CEs care for sick older people without their dual role being recognized in the workplace. Socioculturally, the experience of CEs is complicated by filial piety, the social value of reverence, which significantly constrains the attitudes and behaviors of the caregivers since Hong Kong households emphasize moral obligation in a strictly hierarchical sense based on the recognition of aid and care given to older adults [30].

The salient economic and sociocultural context of Hong Kong is further complicated by Hong Kong's aging population, which is expected to rise sharply from 17% in 2016 to 37% by 2066 [31]. Similarly, the older people dependency ratio is expected to increase from 5:1 in 2015 to 1.8:1 by 2064 [32]. The older people dependency ratio is attributed to the high prevalence of chronic disease in the older population; about 65% of the older adults in Hong Kong have a chronic condition, while one-third have at least two chronic conditions [33]. Furthermore, Hong Kong has the fourth highest cost of living among cities globally [34]. As a result, about 59.6% of the Hong Kong population is significantly engaged in the workforce [35]. These population trends indicate a high need for CFWP in Hong Kong and underscores the urgent need to scale up CFWP in Hong Kong. This need also emerged in our recent study in which 7% of CEs in Hong Kong caring for people with Alzheimer disease had signs of probable clinical depression and 10% possible mild depression [36]. Furthermore, the responses of this study's participants indicated that CFWPs can directly improve their mental well-being and organizational performance [36].

Study Design

This study was envisioned as theory-led research, considering both conventional theories in the field of CEs study and additional selected theories fitting into the context. Spillover theory and IVT were selected to guide and inform this study's design, perspectives, and interpretative lens.

Qualitative case studies are particularly appropriate when contextual conditions believed to be relevant to the phenomenon are under investigation and when boundaries between the phenomenon and context are unclear [37]. Given the absence of CFWP in Hong Kong, the contextual conditions the CEs are subjected to are relevant to CFWP because the potential construction of CFWP is firmly based on the context at hand. Therefore, a qualitative case study methodology involving the three core stakeholders of CFWP, that is, CEs, HPs, and ERs, was used in this study. Individual in-depth semistructured face-to-face interviews were conducted to explore CEs' experiences and needs while incorporating other stakeholders' perceptions, which is well served by a naturalistic, qualitative inquiry and triangulation of perspectives [38]. The data from the interviews were subsequently coded for thematic data analysis.

Hence, the study design adopted for this research allows the issues surrounding CFWP in Hong Kong to be explored via the lenses of multiple stakeholders who may harbor diverse and potentially contrasting perspectives. This method is valuable for health science research in developing interventions (ie, CFWPs) because of its flexibility and rigor in studying complex phenomena using several data sources [37].

Participants

The study participants included three groups: CEs caring for older adults, ERs such as company management personnel, and HPs providing ancillary services for CEs. The following inclusion criteria were used for recruiting CEs: residents of Hong Kong, employed full-time, concurrently taking up a

caregiver role for either their parents or spouses aged older than 60 years, able to provide consent, and that participation would not adversely affect their health or well-being. The eligibility criteria for ER were business owners or executives of a company with at least 10 employees and at least 5 years of management experience; for HP, licensed individuals who provided health care services and had at least 5 years of experience related to caregiver service were eligible.

As the time frame for data collection was limited, a purposive sample was used to select participants who already had some CFWPs-related background knowledge or experience. Potential participants were initially identified by referrals from a scholar in a local research institute specializing in caregiver welfare, followed by a review of participants' prior participation in policy forums and public campaign platforms. This process ensured that the participants were familiarized with the concepts of CE status and the relevant rights and burdens in Hong Kong. Additional participants were identified through the snowballing technique.

Study Instrument

The semistructured in-depth interviews were conducted using a discussion guide consisting of 17 questions (Multimedia Appendix 1) related to three aspects: (1) respondents' personal experience with CEs, (2) their attitudes and preferences for policies, and (3) their perceptions of a caregiver-friendly workplace. The discussion guide was developed based on (1) a thorough literature review, (2) the master's thesis of the first author [39], and (3) the recommendation of two expert panels of researchers specializing in the field of caregivers (ELW from the Chinese University of Hong Kong and XB from the Hong Kong Polytechnic University).

The interview guide was adapted for each participant type. Thus, three sets of interview guides directed the in-depth individual interviews for the three groups of participants. The interview guide for the CEs consisted of a checklist inquiring about the basic demographics of the CEs and their care recipients, followed by open-ended questions about their personal experience on the caregiving journey, encounters on their dual roles, and their interactions with regulatory frames and other actors within the workplace context. The interview guide for ERs started with a checklist to establish basic information about the company, followed by questions about their perspective on the current situation and potential for CFWP. For HPs, the guide outlined their observations and perspectives on CE's experience and CFWP.

Interview Process

The initial interviews were conducted face-to-face in informants' offices but later switched to videoconferencing due to the COVID-19 pandemic. All interviews were audio recorded. Interviews ranged from 20 to 70 minutes. The interviews were conducted in Cantonese (Hong Kong dialect) and then translated and transcribed into English verbatim.

The participant's written consent and demographic information comprising their age, sex, working position, and marital status were obtained before the interview. During the interview, conversation starters were used to open up the participants so they would feel more at ease and provide the most candid response. Prompts were used to elicit more specific and detailed responses to fully explore their experience and perception [40,41].

Thematic Framework Analysis

Thematic framework analysis based on the conceptual guiding framework was adopted. Data coding and analysis were independently conducted by two researchers (MML and XB) to ensure the consistency and reliability of the translation from Chinese to English.

The five stages of the analytical process comprised familiarization, identification of a thematic framework, indexing, charting, and mapping and interpretation [42]. The MAXQDA 2018 analysis tool (VERBI GmbH) was used to index all transcribed data. A total of 209 indexes were charted into 4 parent themes and 65 subthemes. The last stage, mapping and interpretation, was used to gauge the prominence of key themes and subthemes across the full list of participants. While the guiding framework informed the specification of questions and the main categories or themes of interest, the themes specified and explicated by this inquiry reflect participants' views. In other words, any strategy or recommendations emerging from this research echo the participants' attitudes, beliefs, and values [43]. Data credibility was established by obtaining validation from some participants on the accuracy of how their experiences were registered in the form of interview transcripts. Some participants were also invited to review the interpretations to ensure their beliefs were accurately represented, thereby minimizing bias.

Ethical Considerations

Research ethics approval was granted by the Imperial College Research Ethics Committee and The Hong Kong Polytechnic University Department of Applied Social Science Research Committee on June 6, 2018 (reference 18IC4581) and May 1, 2018 (reference HSEARS20180413002), respectively. Written consent was obtained from interview informants after the purpose of the study, and the plan for data confidentiality, analysis, storage, and dissemination was explained to them. All data were anonymized and kept in password-protected folders accessible only to the project supervisor and the research student. Participants were not provided monetary or material compensation for participating in this study.

Results

Participant Demographics

A total of 15 participants were interviewed between May and June 2018. The CEs (n=9; CE1-9) comprised 4 female and 5 male participants aged 20-70 years from four industries (banking or finance, technology, service, and education), with caregiving experience ranging from 0.5 to 13 years (Table 1). Three of the CEs had previously quit jobs because of caregiving roles. The major conditions leading to the caregiving roles were depression, dementia, physical disabilities, and tumors.

The HPs (n=3; HP10-12) comprised a geriatrician (HP10) who has been engaged in geriatric medicine for over 35 years in

private and public practice and two social workers (HP11 and HP12) with over 5 years of experience in different contexts, such as crisis intervention, end-of-life care management, and daily support services (Table 2).

The company personnel (n=3; ER13-15) were drawn from big, medium, and small enterprises engaging in retail, law, and technology, respectively (Table 3). Only ER15 works in a company with a Family-Friendly Employers Scheme by the Hong Kong Government [44].

Table 1. Characteristics of CEs[a] included in this study.

Informant code	Sex	Age group (years)	Industry (seniority)	Condition of the care recipients	Experience as a caregiver (years)	Ever quit a job because of the caregiving role
CE1	F[b]	21-30	Bank (junior)	Brain tumor	5	No
CE2	F	51-60	Retired; bank (senior)	Depression	6	No
CE3	F	21-30	Accounting (middle)	Lung cancer	2	Yes
CE4	F	41-50	Technology	Depression	10	No
CE5	F	31-40	Primary school teacher	Joint disorder	2	No
CE6	M[c]	51-60	Security guard	Physical disabilities	0.5	No
CE7	M	Older than 60	Retired; bank (senior)	Dementia	13	Yes
CE8	M	41-50	Servicing industry (middle)	Dementia	10	No
CE9	M	41-50	Secondary school teacher	Stroke	8	Yes

[a]CE: caregiver-employees.

[b]F: female.

[c]M: male.

Table 2. Characteristics of HPs[a] included in this study.

Informant code	Profession	Years of practice
HP10	Geriatrician	>35
HP11	Social worker and gerontologist; specializes in end-of-life care	>5
HP12	Social worker; specializes in caregiver services	>5

[a]HP: health care professional.

Table 3. Characteristics of ERs[a] included in this study.

Informant code	Company size	Industry	Position	Awardee of family-friendly employers
ER13	200	Legal	Owner	No
ER14	10	Technology	Senior Manager	No
ER15	2000	Retails	Director, Human Resources	Yes

[a]ER: employer.

Current Overview

Interviews with all three CFWP stakeholder groups indicated the lack of formal workplace policy directly addressing CE issues. CEs further highlighted that any accommodation to their needs was made by management on a case-by-case, informal, and discretionary basis. Only two participants reported some form of formal workplace support for CEs. These family-friendly policies included a home office program and monthly 2-hour early leave for CE1 and CE4. These policies applied to all employees where CE1 and CE4 worked. Both CEs perceived these policies positively, exemplified by the remarks:

> *It makes a difference. It has reduced stress drastically. It is much easier to manage time, physical health (tiredness), and mental and emotional needs.* [CE4]

Thematic Analysis

Overview

The thematic analysis yielded 65 codes (Multimedia Appendix 2) categorized into four main themes: (1) lacking leadership, (2) cultural factors, (3) role struggles, and (4) inadequate support.

Lacking Leadership

A lack of guidance in laws or social policies for realizing CFWP in Hong Kong was evident during the interviews. All ER participants perceive that companies lack the resources, knowledge, and experience to support CEs. They consider government leadership essential, such as subsidies, guiding policy, and technical support. All three ER informants expressed willingness to adopt CFWP if the government takes the lead first. For example, one ER stated:

> *Government taking the lead is important. As long as there are some initiatives in the Employment Ordinance, corporates will initiate to follow...Citing paternal leave is a good example...How can he be part of the family throughout the journey? If considered, you could be surprised what the corporates would offer.* [ER15]

Within the organizational leadership, many operational factors and competing priorities influence ERs' decisions in adopting CFWP in Hong Kong, such as cost, fairness, potential uptake, and inadequate resources. ER13 reflected that caregiving roles and mental illnesses are stigmatized in Hong Kong, limiting potential service utilization. Smaller companies are hesitant to adopt CFWP officially for the potential cost implications and further responsibilities it might entail. ER14 remarked that formalization may unleash other obligations. ER14 and ER15 admitted they have other priorities, such as childcare, over CFWP. For example, ER15 highlighted, "Corporates might perceive not all older adults need help, but all children need care." HP10 confirmed that children-friendly work arrangements are common but not CFWP in Hong Kong.

Cultural Factors

CFWP aims to promote holistic integration of work and life by enabling CEs to balance their roles as employees and caregivers. However, some observations made by participants seem to suggest that the fundamental tenets of CFWP may be at odds with cultural and workplace values in the Hong Kong setting. The concept of work-life separation, rather than work-life balance, is highly prevalent in Hong Kong. A total of 6 (67%) out of 9 CEs indicated that it is an established norm in Hong Kong to separate personal or private concerns from work, if not altogether, rendering them secondary to work concerns. In other words, personal issues should not be imposed on workplaces. Failure to detach from personal stress at work could entail a reproach from management. For instance, CE7 recalled that his manager remarked, "Don't bring this burden to work. Once you clock in, you better not have this burden in your mind."

Four CE participants stated that the prevalent corporate attitude in Hong Kong prioritizes business interests over employees' well-being. The participants highlighted a culture of assumption of total work commitment from employees that leaves little room for considering CEs' concerns. They further intimated that this culture makes it difficult for CEs to raise the issue of caring for loved ones with their managers. HP10 shared this sentiment rather strongly while stating, "Employers expect slavery, working 10 hours, better not to go home, not taking leave, not getting sick, without seniors at home, no marriage, no relationship, and no children."

This attitude of management caused acute stress for employees and potential conflict with management when urgent health issues of the family care recipients need to be dealt with while the CE is at work. At these times, CEs find themselves in a dilemma of multiple roles and uncertainties. For example, one CE stated:

> *It was so stressful every time my mother called when I was at work...I was so nervous. Because of the long-term mental stress, I have hypertension.* [CE7]

All CEs lamented a general lack of compassion from their ERs and colleagues. CEs must take formal leave because "caregiving roles occur outside work hours." As a result of the unsympathetic attitude, 6 CEs chose not to disclose their CE status and remained silent, while others became more stressed or decided to quit. CE5 described her organization's management as so apathetic that she chose not to disclose her chronic condition as well. She regarded her role in the organization as "replaceable." Therefore, letting management know her caregiving role was not an option. She felt that she had no power or involvement in the decision-making process concerning her disease or her caregiving role because of the hierarchy of her organization. CE5 and HP11 attributed this passivity from the ERs to the assumption that the government should take full responsibility for supporting CEs.

CEs also spoke of colleagues' indifference to their circumstances. Fear of gossip limits CEs from disclosing their circumstances. Disclosing could be perceived as an excuse to do less work, as exemplified by a statement:

> *Trust is important, so information about me will not be spread around...I don't want people to see it as an excuse [weakness]. I don't want people to see me using my mother's condition as an excuse to do less work.* [CE3]

CE6, a security guard, believes that disclosure could dangerously affect his position in the organization. Conversely, CE4, a senior manager, expressed no hesitation or concerns regarding disclosure of his CE status. This indicates that seniority in the corporate hierarchy influences how empowered CEs may feel about disclosing their status.

The notion of the primacy of work over personal concerns is also reflected in CEs' behavior. Most CEs took it upon themselves to overcome additional burdens at work and were either not keen or could not seek help and share burdens with others. They expressed high expectations of themselves as employees and caregivers, compounding the stress. One HP noted:

> *They (CEs) will just impose the problems on themselves. They choose not to take absences but work until midnight to care for the older adults.* [HP10]

This work ethic exacerbates the stress experienced by the CEs.

Role Struggle

All CEs expressed experiencing struggles navigating the different roles as employees, caregivers, parents, partners, and friends. They felt the urge to reprioritize values in life as they

often take on a reverse role, from a child to a guardian for their frail parents. CEs reported that the caregiving role influenced their behaviors and decisions, like spending patterns or quitting a job. For instance, one CE stated:

> *I didn't want to abuse the system or the job and wanted to prioritize my students' learning experience, and that's why I decided to quit.* [CE9, teacher]

CEs sacrificed their time to undertake the caregiver role at the expense of self-care, potentially causing a loss of personal identity and difficulties connecting with themselves and others. As CE1 stressed, "The biggest change is the loss of friends and my personal time. All my time was devoted to family needs." In the case of CE7, he was abandoned by his wife after his full-time caregiving role forced him to quit his job. Thus, reprioritization of values could wreak havoc on family integrity. In addition, reprioritization may result in unmet psychological and emotional needs, introducing an identity struggle that further overwhelms the CEs. In this regard, CE7 remarked, "I need to look at myself as another person to live up to my caregiving role."

It is difficult to strike a balance between caregiver and employee roles. The intensity and stress of these roles are inseparable and nonexclusive, which may lead to a downward spiral if no support is available. New CEs may be more prone to this balancing problem because they lack institutional support and resources to prepare for these dual roles. For instance, one CE stated:

> *It was very difficult to adjust at the beginning...Letting go of work stress and switching to caregiver mode was very difficult.* [CE1]

Conversely, work could also be a protective factor for CEs to take on caregiving roles. Engagement at work can alleviate the caregivers' stress if balance can be attained. This was best exemplified by the statement:

> *If the caregiving role is 24/7, you will lose the meaning and make caregiving an obligation. It will become a burden. Especially in Hong Kong, where living space is so small. If you only take on the caregiving role, it can feel like being trapped in a cage.* [CE8]

The unavoidable familial tension—reported as "unresolved childhood emotions"—further rarifies the role struggle. HP12 spoke about how the inner child of the caregivers can be a vulnerability: "Since the care recipients are the parents who have shaped a large part of the caregivers' lives, a single gesture or action could trigger emotional stress from trauma as far [back] as those from childhood." CE5 exemplified inner child struggle as well. She reflected that she found it unfair that she had to care for her father, who had done little to support the family during her childhood. Even though she has become financially independent, CE5 would readily associate the overwhelming stress of the caregiving role with the powerlessness and lack of confidence in her child-self because of the wounds caused by her father. These emotions could intensify the present role struggle.

Inadequate Support

CEs exhibited a feeling of hopelessness mainly because of the uncertainties they faced. This hopelessness is exacerbated by the minimal assistance and support received, as explained by one HP:

> *Hopelessness...because chronic illnesses have no foreseeable end. The uncertainty of knowing how long the suffering will last causes hopelessness. The caregivers' task lists don't end either. It never ends. It is endless.* [HP11]

There is limited information to prepare and enable caregivers to live up to their roles. Besides limited information from doctors, there is a lack of reader-friendly, timely, and high-quality information. HP 10 remarked, "There is a serious need for information support...Hearsay is commonly found but useless." All CEs were frustrated in finding the quality information they needed to live up to their roles, including those related to social welfare support and handling the needs of care recipients. Most CEs had little idea about available social welfare support resources because of limited public awareness and promotion programs. In this context, one CE stated:

> *The hospital didn't actively promote them to us. And we didn't expect there is this kind of thing. It was difficult to find appropriate information. Doctors at public hospitals would not have the time to explain—only standard treatment procedures. But they didn't talk about the side effects or things about daily care to know. I was quite lost.* [CE3]

Most CEs perceived the information provided by social welfare support as inadequate. One of the social worker informants confirmed that due to potential conflicts of interest in recommending a specific service provider to caregivers, it is normal for caregivers to receive a long list of providers without recommendations, which can be confusing. The inadequate quality of information and the lack of promotional programs cause an information gap that makes CEs' caregiving journey challenging.

Accessing social welfare support is limited because of the high qualification threshold. Most cases are in the middle of the scale and ineligible for financial or physical support like respite care or home care services. In addition, the resources designated for caregivers are generally meager. One CE responded:

> *The social worker has only 1% of her work duties devoted to handling cases...she made decisions on my behalf without telling me about the rationale and the context.* [CE9]

Such negative experiences further discourage CEs from seeking social welfare support. ERs noted the need to match the services of providers with users, particular conditions they are facing, and the stage in the caring process. HP13 reported that CEs' characteristics influence service quality and satisfaction. Matching is crucial in the first stage to ensure good quality, according to ER9.

Health care providers agreed that social welfare support is inadequate because of fragmentation and lack of crisis

intervention. Crisis intervention services are available only to hospitalized cases. The chances of accessing them after discharge are minimal. The care system is fragmented and superficial because social workers have other challenging roles besides carer services. Social workers might chase the number of cases under their purview instead of going deeper individually by enhancing the caregivers' well-being or counseling them.

Besides limited resources and services, operations and priority matrix do not favor caregivers' referrals for social support, as exemplified by the following statement by HP10: "Relying exclusively on their doctors is not enough when these doctors are extremely busy." CE3 further reflected that health care providers' attitudes often lacked empathy and professionalism, resulting in disappointment and frustration. The lack of empathy or indifferent attitude of health care practitioners intensifies CEs' powerlessness. In this regard, CE3 stated, "Since time and resources are scarce in the public hospital, they wouldn't prepare you mentally for the situation. They would just tell me the symptoms and options and seek our consent for surgery immediately. There was no time allowed for our consideration. Everything was mechanical."

There are also substantial barriers to accessing professional mental health services for caregivers. For example, waiting time is considerably long, and according to HP12, "only serious cases, such as suicidal attempts, might be considered for the clinical mental health services."

Discussion

Principal Findings

This study explored the CFWP-related views and experiences of multiple stakeholders (CEs, ERs, and HPs) in Hong Kong, guided by a theoretical framework that oriented the interest toward exploring CE experiences in depth (IVT and spillover theory) and exploring the potentials of CWFP as new ideas within the broader context (PARIHS framework). Principal findings were broadly organized under two context-related ("lacking leadership" and "unfavorable culture") and two facilitation-related ("role struggle" and "inadequate support") themes.

Regarding CFWP, Hong Kong lags behind other aging Asian societies, such as Japan and Taiwan, which have already adopted 93 [45] and 21 [46] leave days for CEs, respectively. Lack of clear leadership and confusion in public policy roles and guidance was highlighted as one of the primary reasons why explicit CFWPs are absent in Hong Kong, despite the urgent need for CEs and the apparent willingness of ERs to adopt CFWP. The government taking the lead in establishing guiding frameworks for CFWP, both in law and administration, appears to be the critical missing link. Given that Hong Kong's health care system bears colonial inheritance [47], the United Kingdom's Care Act could prove instructional. The UK Care Act was implemented in 2014 to protect the rights of caregivers. The law provides provisions for the local authorities to identify CEs' needs by assessing private companies [48]. These provisions allow the UK government to strategically partner with private companies and provide support and services for ERs to achieve caregiver-friendly working environments [49]. Similar legal protection and policies adapted to the local context will be instrumental in widespread CFWP adoption among Hong Kong companies.

Regarding organizational leadership, welfare ideology and operational concerns are the two factors hindering the formulation and uptake of CFWP. Hong Kong's long-standing universal welfare ideology, described as distorted toward the government presumably providing for welfare in aging and antipoverty programs [50], fuels companies' denial of their corporate responsibility in addressing the needs of the CEs. This study also shows that operational factors, such as cost, potential uptake, and inadequate resources, prevent CFWP from gaining prominence in the organizational leadership agenda. However, the absence of CFWP is far more costly in the long run due to increased absenteeism, reduced work productivity, increased turnover, and work disruptions [51]. Sustainable long-term organizational growth would require well-considered organizational leadership that addresses this substantive issue.

In addition, local cultural factors were also identified as hindrances to CFWP adoption, such as issues concerning work-life separation, total work commitment, and lack of compassion from management. The Confucian work ethic of hard work, perseverance, and patience is also deeply embedded among Hong Kongers [52]. The sentiments and behavior reflected in our study affirm the prevalence of these values and further indicate that these may not always be compatible with the underlying principles of CFWP. The associated taboos and risks of speaking up, as perceived and experienced by CEs, further demonstrate the assertions of IVT that explain the risks and inappropriateness of speaking up.

Older adults may have diverse and unexpected health care needs that are not always associated with chronological age [53]. Thus, adopting a caregiver-friendly workplace culture conducive to compassion and acceptance of CE's circumstances is necessary. Such cultures include supportive management, a trusting environment, and establishing top-to-bottom leadership [54]. An empathic and caregiver-friendly culture would potentially encourage CEs to identify themselves (rather than implicit voice spaces), and incorporating flexibility and acceptance into business management and work culture enhances employee commitment [55].

Regarding role struggles, our findings indicate that there may be positive and negative spillovers between work and family, upholding the constructs of spillover theory. For some CEs, work can serve as a break from caregiving roles, and hence, be a protective factor, while the high burden of dual roles can force some CEs to quit work, sacrifice self-care, or even get abandoned by family members. For instance, a study focusing on the well-being of Hong Kong male caregivers reported that work has protective effects by improving caregivers' resilience and overall well-being [56]. Therefore, role-balancing is key because it rewards CEs with a sense of satisfaction and fulfillment and also avoids any ramifications by giving up paid employment for caregiving roles [57]. CFWP, which promotes and engages CEs to balance work and life, is critical for positively managing this spillover effect. These findings are

consistent with existing literature that CFWP reduces absenteeism and sick leave rates and increases employees' productivity, loyalty, engagement, and morale [58-60].

The inadequacy of support was amply reflected by CEs and HPs alike. Even though family caregiving has become more demanding, complicated, and longer-lasting than in the past, caregivers are often underprepared [61]. This study further highlights that CPWP is most needed by CEs who have just commenced the caregiving role and experience acute crisis because of little preparation and knowledge about their new role. This, in turn, emphasizes how support for new caregivers is critical for the short and long-term well-being of both CEs and their care recipients. Caregiver preparation and education are vital in reducing psychological stress and coping with the situation better [62-64]. Companies can bridge the information gap and facilitate matching according to their unique demographics and the needs of caregivers. CFWP is, therefore, imperative to help new CEs navigate their caregiving journey.

Previously, caregiving tasks in Hong Kong could be shared among siblings when family sizes were larger. However, nuclear families have become the norm in Hong Kong, with an average household size of 2.8 members [65]. Future generations are expected to have even fewer siblings sharing caregiving responsibilities, which is bound to compound and intensify the burden and sense of powerlessness. Therefore, effective CFWP is needed to ensure support for income earners to avoid the myriad adverse effects on individuals, society, and businesses.

Limitations

The major limitation of this study is that we predominantly recruited CEs engaged in full-time work, which may not fully reflect the views of CEs who are part-time or unemployed. In addition, CE participants in this study were well primed to reflect on these issues as they were already actively engaged in policy consultations for CFWP, limiting the generalizability of our findings.

Conclusions

This study is the first inquiry into the experiences of CEs and factors that may influence the adoption of CFWP in Hong Kong. The qualitative methodology and the purposively identified participants have provided a glimpse into the lives and tribulations of this otherwise invisible but growing population group. Our findings strongly indicate that Hong Kong's current workplace policy frameworks fall short of meeting the immediate needs of CEs and the long-term interests of companies and society. We urge actors to explore and adopt potential CWFP for Hong Kong, considering the complexity of factors explored in this study, including unique cultural dynamics and structural factors related to the aging population and ever-increasing health care burden. Future research should identify CEs who need the most support and pinning down the most optimal forms of regional law and organizational policy.

Acknowledgments

This manuscript did not receive any research funding.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available for privacy reasons but are available from the corresponding author upon reasonable request.

Authors' Contributions

MML, HT, XB, NCY, and CF built the study concepts and constructed the study design. MML conducted the interviews, wrote the manuscript, and provided data for all tables. MML, HT, and CF analyzed and interpreted the results. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The interview guide.
[DOCX File , 18 KB - ijmr_v14i1e58528_app1.docx]

Multimedia Appendix 2
Emergent themes, subthemes, and codes from thematic framework analysis.
[DOCX File , 14 KB - ijmr_v14i1e58528_app2.docx]

References

1. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-01-31]

2. Grinin L, Grinin A, Korotayev A. Global aging: an integral problem of the future. How to turn a problem into a development driver? In: Sadovnichy V, Akaev A, Ilyin I, Malkov S, Grinin L, Korotayev A, editors. Reconsidering the Limits to Growth. Cham: Springer International Publishing; 2023:117-135.
3. Boyle D. Caregiving within the context of elder care. In: Informal Caregivers: From Hidden Heroes to Integral Part of Care. Cham, Switzerland: Springer International Publishing; 2023:33-67.
4. Sawik B, Tobis S, Baum E, Suwalska A, Kropińska S, Stachnik K, et al. Robots for elderly care: review, multi-criteria optimization model and qualitative case study. Healthcare (Basel) 2023;11(9):1286 [FREE Full text] [doi: 10.3390/healthcare11091286] [Medline: 37174828]
5. Robison J, Shugrue N, Fortinsky RH, Gruman C. Long-term supports and services planning for the future: implications from a statewide survey of baby boomers and older adults. Gerontologist 2014;54(2):297-313. [doi: 10.1093/geront/gnt094] [Medline: 23990621]
6. Del-Pino-Casado R, Priego-Cubero E, López-Martínez C, Orgeta V. Subjective caregiver burden and anxiety in informal caregivers: a systematic review and meta-analysis. PLoS One 2021;16(3):e0247143 [FREE Full text] [doi: 10.1371/journal.pone.0247143] [Medline: 33647035]
7. Dich N, Lange T, Head J, Rod NH. Work stress, caregiving, and allostatic load: prospective results from the Whitehall II cohort study. Psychosom Med 2015;77(5):539-547 [FREE Full text] [doi: 10.1097/PSY.0000000000000191] [Medline: 25984826]
8. Bauer JM, Sousa-Poza A. Impacts of informal caregiving on caregiver employment, health, and family. Popul Ageing 2015;8(3):113-145. [doi: 10.1007/s12062-015-9116-0]
9. Yeandle S, Bennett C, Buckner L, Shipton L, Suokas A. Who Cares Wins: The Social and Business Benefits of Supporting Working Carers. London, United Kingdom: CarersUK; 2006:2016.
10. Juratovac E, Zauszniewski JA. Full-time employed and a family caregiver: a profile of women's workload, effort, and health. Womens Health Issues 2014;24(2):e187-e196. [doi: 10.1016/j.whi.2014.01.004] [Medline: 24630422]
11. Reinhard SC, Given B, Petlick NH, Bemis A. Supporting family caregivers in providing care. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality (US); 2008.
12. Matos K, Galinsky E, Bond JT. 2016 National study of employers. Society for human resource management. 2017. URL: https://cdn.sanity.io/files/ow8usu72/production/d73a7246cc3a3fef4ad2ece1e3d5aa4eaec2f263.pdf [accessed 2021-07-05]
13. Ireson R, Sethi B, Williams A. Availability of caregiver-friendly workplace policies (CFWPs): an international scoping review. Health Soc Care Community 2018;26(1):e1-e14 [FREE Full text] [doi: 10.1111/hsc.12347] [Medline: 27071368]
14. Vakkayil J, Torre ED, Giangreco A. “It's not how it looks!” exploring managerial perspectives on employee wellbeing. Eur Manage J 2017;35(4):548-562. [doi: 10.1016/j.emj.2016.12.002]
15. Ernst KE, Kalliath T, Kalliath T, Kalliath P. Achieving employee wellbeing in a changing work environment. Int J Manpower 2012;33(7):738-753. [doi: 10.1108/01437721211268294]
16. van de Voorde K, Paauwe J, van Veldhoven M. Employee well‐being and the HRM–organizational performance relationship: a review of quantitative studies. Int J Manage Rev 2011;14(4):391-407. [doi: 10.1111/j.1468-2370.2011.00322.x]
17. Lorenz F, Whittaker L, Tazzeo J, Williams AM. Availability of caregiver-friendly workplace policies: an international scoping review follow-up study. IJWHM 2021;14(4):459-476. [doi: 10.1108/ijwhm-10-2019-0136]
18. Ding R, Dardas A, Wang L, Williams A. Evaluation of a caregiver-friendly workplace program intervention on the health of full-time caregiver employees: a time series analysis of intervention effects. J Occup Environ Med 2020;62(10):e548-e558 [FREE Full text] [doi: 10.1097/JOM.0000000000001980] [Medline: 32769780]
19. Mofidi A, Tompa E, Williams A, Yazdani A, Lero D, Mortazavi SB. Impact of a caregiver-friendly workplace policies intervention: a prospective economic evaluation. J Occup Environ Med 2019;61(6):461-468. [doi: 10.1097/JOM.0000000000001564] [Medline: 31166280]
20. Sethi B, Williams A, Ireson R. Supporting caregiver employees: managers’ perspective in Canada. IJWHM 2017;10(1):25-41. [doi: 10.1108/ijwhm-03-2016-0016]
21. Rycroft-Malone J. The PARIHS framework—a framework for guiding the implementation of evidence-based practice. J Nurs Care Qual 2004;19(4):297-304. [doi: 10.1097/00001786-200410000-00002] [Medline: 15535533]
22. Sirgy MJ, Efraty D, Siegel P, Lee DJ. A new measure of quality of work life (QWL) based on need satisfaction and spillover theories. Social Indic Res 2001;55(3):241-302. [doi: 10.1023/a:1010986923468]
23. Knoll M, Neves P, Schyns B, Meyer B. A multi-level approach to direct and indirect relationships between organizational voice climate, team manager openness, implicit voice theories, and silence. Appl Psychol 2020;70(2):606-642. [doi: 10.1111/apps.12242]
24. Hill E, Ferris M, Märtinson V. Does it matter where you work? A comparison of how three work venues (traditional office, virtual office, and home office) influence aspects of work and personal/family life. J Vocational Behav 2003;63(2):220-241. [doi: 10.1016/s0001-8791(03)00042-3]
25. Milliken FJ, Morrison EW, Hewlin PF. An exploratory study of employee silence: issues that employees don’t communicate upward and why*. J Manage Stud 2003;40(6):1453-1476. [doi: 10.1111/1467-6486.00387]

26. Chan MK. Stability and prosperity in Hong Kong: the twilight of laissez-faire colonialism?—A review article. J Asian Stud 2011;42(3):589-598. [doi: 10.2307/2055519]
27. Chiu SWK, Siu KYK. Hong Kong as an economic miracle? The myth of laissez-faire and industrialization. In: Hong Kong Society. Singapore: Springer Nature Singapore; 2022:137-163.
28. Tang G, Yuen RHY. Hong Kong as the 'Neoliberal Exception' of China: transformation of Hong Kong citizenship before and after the transfer of sovereignty. J Chin Polit Sci 2016;21(4):469-484. [doi: 10.1007/s11366-016-9438-7]
29. Peck J. Milton's paradise: situating Hong Kong in neoliberal lore. JLPE 2021;1(2):189-192. [doi: 10.5070/lp61251592]
30. Ng WI, Che SL, Li X, Zhu M. Association of filial attitude, filial behavior and death literacy: implications for development of death system in Guangdong-Hong Kong-Macao Greater Bay area of China. BMC Public Health 2024;24(1):721 [FREE Full text] [doi: 10.1186/s12889-024-18197-3] [Medline: 38448863]
31. Hong Kong population projections. Census and Statistics Department (Hong Kong Special Administrative Region Government). 2017. URL: https://www.censtatd.gov.hk/en/data/stat_report/product/FA100061/att/B71710FA2017XXXXB0100.pdf [accessed 2021-07-05]
32. Population policy. Chief Secretary for Administration's Office (Hong Kong Special Administrative Region Government). 2015. URL: https://gia.info.gov.hk/general/201501/15/P201501150880_0880_140608.pdf [accessed 2021-07-05]
33. Wong SYS, Zou D, Chung RY, Sit RW, Zhang D, Chan D, et al. Regular source of care for the elderly: a cross-national comparative study of Hong Kong With 11 developed countries. J Am Med Dir Assoc 2017;18(9):807.e1-807.e8. [doi: 10.1016/j.jamda.2017.05.009] [Medline: 28647578]
34. Worldwide cost of living 2022: how soaring inflation has affected prices globally. The Economist Intelligence Unit Limited. 2022. URL: https://www.eiu.com/n/wp-content/uploads/2022/11/Worldwide-Cost-of-Living-2022.pdf [accessed 2024-02-01]
35. Labour force 2021. Census and Statistics Department (Hong Kong Special Administrative Region Government). URL: https://www.censtatd.gov.hk/en/web_table.html?id=7 [accessed 2021-07-05]
36. Lee MMS, Yeoh EK, Wang K, Wong ELY. Caregiver employees' mental well-being in Hong Kong. Healthcare (Basel) 2024;12(10):1013 [FREE Full text] [doi: 10.3390/healthcare12101013] [Medline: 38786422]
37. Yin RK. Case Study Research Design and Methods. Thousand Oaks, CA: Sage; 2014:282.
38. Flick U. Doing Triangulation and Mixed Methods. London, United Kingdom: SAGE Publications Ltd; 2018.
39. Lee M, French C, Bai X, Yeung C, Taddese H. A Case Study of Caregiver-Friendly Workplace Policy in Hong Kong [master's thesis]. London, UK: Imperial College London; Jun 2018:1-20.
40. Service RW. Techniques and procedures for developing grounded theory. In: Corbin J, Strauss A, editors. Organizational Research Methods. Thousand Oaks, CA: SAGE Publications, Inc; 2008:614-617.
41. Crewell JW, Crewell JD. Research Design—Qualitative, Quantitative, and Mixed Methods Approaches. Washington, DC: Sage Publication; 2017.
42. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Byrman AG, Burgess R, editors. Analyzing Qualitative Data. London, United Kingdom: Routledge; 1994:173-194.
43. Srivastava A, Thomson SB. Framework analysis: a qualitative methodology for applied policy research. Journal of Administration and Governance 2009;4:72-79 [FREE Full text]
44. Family-friendly employers award scheme 2017/18. Family Council. 2022. URL: https://www.familycouncil.gov.hk/en/Employment/employers2017.html [accessed 2022-03-14]
45. Nishimoto M. Choices of leave when caring for family members: What is the best system for balancing family care with employment. Jpn Labor Rev 2017;14(1):45-67 [FREE Full text]
46. Family caregivers. Taiwan Association of Family Caregivers. 2022. URL: https://www.familycare.org.tw/policy/901 [accessed 2022-04-18]
47. Kong X, Yang Y, Gao J, Guan J, Liu Y, Wang R, et al. Overview of the health care system in Hong Kong and its referential significance to mainland China. J Chin Med Assoc 2015;78(10):569-573 [FREE Full text] [doi: 10.1016/j.jcma.2015.02.006] [Medline: 26144022]
48. Ahmed R. Care act FAQ. Carers UK. 2022. URL: https://www.carersuk.org/help-and-advice/guides-and-tools/your-guide-to-the-care-act-england/ [accessed 2022-03-10]
49. Hockin H. Partner with us. Carers UK. 2022. URL: https://www.carersuk.org/how-you-can-help/partner-with-us [accessed 2022-03-10]
50. Yu A. The failure of the welfare ideology and system in Hong Kong: a historical perspective. Hum Aff 2021;31(1):99-108. [doi: 10.1515/humaff-2021-0009]
51. Keating NC, Fast JE, Lero DS, Lucas SJ, Eales J. A taxonomy of the economic costs of family care to adults. J Econ Ageing 2014;3:11-20. [doi: 10.1016/j.jeoa.2014.03.002]
52. Wong D. Stress factors and mental health of carers with relatives suffering from schizophrenia in Hong Kong: implications for culturally sensitive practices. Br J Social Work 2000;30(3):365-382. [doi: 10.1093/bjsw/30.3.365]
53. World report on ageing and health. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/186463 [accessed 2024-02-01]
54. Bernard M, Phillips JE. Working carers of older adults. Commun Work Fam 2007;10(2):139-160. [doi: 10.1080/13668800701270075]

55. Dex S. Families and work in the twenty-first century. 2003. URL: https://www.researchgate.net/publication/239987709_Families_and_work_in_the_twenty-first_century [accessed 2024-02-01]
56. Siu OL. Research on family-friendly employment policies and practices in Hong Kong: implications for work-family interface. In: Lu L, Cooper C, editors. Handbook of Research on Work-Life Balance in Asia. Cheltenham, United Kingdom: Edward Elgar Publishing; 2015.
57. Horrell B, Breheny M, Stephens C. Capability to care and work: when dual roles intersect. Vulnerable Groups Inclusion 2014;5(1):24357. [doi: 10.3402/vgi.v5.24357]
58. Barroso M. EuroWORK case study: work adjustments, Axa, Portugal. Eurofound. 2011. URL: https://www.eurofound.europa.eu/en/case-studies-0 [accessed 2024-02-01]
59. Peters P. EuroWORK case study: awareness raising, Novel Zorg and Wonen, the Netherlands. Eurofound. 2011. URL: https://www.eurofound.europa.eu/en/case-studies-0 [accessed 2024-02-01]
60. Humer Z. EuroWORK case study: awareness raising, Mercator, Slovenia. Eurofound. 2011. URL: https://www.eurofound.europa.eu/en/case-studies-0 [accessed 2024-02-01]
61. Committee on Family Caregiving for Older Adults, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. In: Schulz R, Eden J, editors. Families Caring for an Aging America. Washington, DC: National Academies Press (US); 2016.
62. Schumacher KL, Stewart BJ, Archbold PG, Caparro M, Mutale F, Agrawal S. Effects of are giving demand, mutuality, and preparedness on family caregiver outcomes during cancer treatment. Oncol Nurs Forum 2008;35(1):49-56. [doi: 10.1188/08.ONF.49-56] [Medline: 18192152]
63. Williamson GM, Schulz R. Caring for a family member with cancer: past communal behavior and affective reactions. J Appl Soc Pyschol 2006;25(2):93-116. [doi: 10.1111/j.1559-1816.1995.tb01586.x]
64. Wang YN, Shyu YIL, Tsai WC, Yang PS, Yao G. Exploring conflict between caregiving and work for caregivers of elders with dementia: a cross-sectional, correlational study. J Adv Nurs 2013;69(5):1051-1062. [doi: 10.1111/j.1365-2648.2012.06092.x] [Medline: 22776026]
65. Population and household statistics. Census and Statistics Department (Hong Kong Special Administrative Region Government). 2021. URL: https://gia.info.gov.hk/general/202103/30/P2021033000302_363913_1_1617079632759.pdf [accessed 2024-02-01]

Abbreviations

CE: caregiver-employee
CFWP: caregiver-friendly workplace policy
ER: employer
HP: health care professional
IVT: implicit voice theory
PARIHS: Promoting Action on Research Implementation in Health Services

Edited by T de Azevedo Cardoso; submitted 18.03.24; peer-reviewed by D Huang; comments to author 17.10.24; revised version received 03.11.24; accepted 19.12.24; published 10.02.25.

Please cite as:
Lee MMS, Yeoh EK, Wong ELY, Bai X, Yeung NCY, French C, Taddese H
Perceptions and Experiences of Caregiver-Employees, Employers, and Health Care Professionals With Caregiver-Friendly Workplace Policy in Hong Kong: Thematic Analysis
Interact J Med Res 2025;14:e58528
URL: https://www.i-jmr.org/2025/1/e58528
doi:10.2196/58528
PMID:

Original Paper

Evaluating the Effect of the JUUL2 System With 5 Flavors on Cigarette Smoking and Tobacco Product Use Behaviors Among Adults Who Smoke Cigarettes: 6-Week Actual Use Study

Nicholas I Goldenson[1], PhD; Saul Shiffman[2], PhD; Mark A Sembower[2], MS; Arielle Selya[2], PhD; Steve Pype[2], BS; Ryan A Black[1], PhD

[1]Juul Labs, Inc, Washington, DC, United States

[2]PinneyAssociates, Inc, Pittsburgh, PA, United States

Corresponding Author:
Nicholas I Goldenson, PhD
Juul Labs, Inc
1000 F Street NW
Suite 800
Washington, DC, 20004
United States
Phone: 1 510 501 5964
Email: nicholas.goldenson@juul.com

Abstract

Background: Adults who switch completely from smoking cigarettes to using electronic nicotine delivery systems (ENDS) substantially reduce their exposure to toxicants and carcinogens that are associated with smoking-related diseases.

Objective: This 6-week actual use study—a prospective uncontrolled real-world study designed to evaluate quasi-naturalistic product use—aimed to assess switching behavior among US adults who smoked cigarettes and were provided with JUUL2 ENDS products.

Methods: US adults who smoked cigarettes every day but were predominantly not ready to quit (N=1160; mean age 39.42, SD 11.03 years; 641/1160, 55.26% female participants; 667/1160, 57.5% non-Hispanic White; mean cigarettes per day 14.11, SD 8.96; only 1% [11/1160] planning to stop smoking within 30 days; and 481/1160, 41.47% dual users) were recruited to use JUUL2 ENDS products (18 mg/mL nicotine) in 1 of 5 flavors in real-world environments for 6 weeks. Participants who expressed sufficient interest in using JUUL2 products were enrolled at 24 different consumer research sites across the United States into one of the two following study arms: (1) traditional flavors (Virginia Tobacco and Polar Menthol, 10 sites); or (2) complex flavors (Autumn Tobacco, Summer Menthol, and Ruby Menthol, 14 sites). No instructions regarding JUUL2 product use or cigarette smoking were provided. After a 1-week trial period, participants were provided with their preferred flavor for 6 weeks of ad libitum use (10 pods per week). In total, 6 weekly web-based surveys were used to assess switching (smoking abstinence) and smoking reduction; dependence and respiratory symptoms were assessed at baseline and week 6.

Results: Across the 5 flavor groups at week 6, the rates of complete past-7-day switching away from cigarettes ranged from 38.2% (79/207) to 47.3% (95/201), and 24.3% (55/226) to 33.9% (74/218) of participants reported complete past-30-day switching. Participants who used the 3 menthol-flavored (vs 2 tobacco-flavored) JUUL2 products had significantly higher rates of past-30-day switching at week 6 (odds ratio 1.36, 95% CI 1.04-1.78). Compared to their baseline values when they were smoking, the past-30-day switchers at week 6 had significantly reduced their dependence (mean differences in dependence, cigarettes – JUUL2: 0.57-0.99; $P<.001$) and self-reported frequency of respiratory symptoms ($P<.05$). Among participants who continued to smoke at week 6, 50.9% (59/116) to 62.9% (73/116) reduced their daily cigarette consumption by at least 50% from baseline.

Conclusions: Adoption of JUUL2 ENDS products can likely help substantial proportions of US adults who smoke to switch completely away from cigarettes or meaningfully reduce their cigarette consumption, thereby reducing their dependence on tobacco products and improving their respiratory symptoms.

(Interact J Med Res 2025;14:e60620) doi:10.2196/60620

KEYWORDS

cigarette; smoking; electronic nicotine delivery system; switching; actual use study; real world

Introduction

Background

Evidence demonstrates that adults who switch completely from smoking cigarettes to using electronic nicotine delivery systems (ENDS) substantially reduce their exposure to toxicants and carcinogens that are associated with smoking-related diseases [1-4]. Accordingly, the primary public health benefit of ENDS is derived from facilitating switching behavior among adults who smoke.

For ENDS products that are currently marketed in the United States, observational studies and randomized trials have assessed switching behavior among adults who smoke [5-9]. However, for premarket ENDS products (ie, those not currently marketed in the United States), evaluating naturalistic product use behaviors and transitions over time in real-world settings requires providing the premarket product to adults who smoke cigarettes. Actual use studies—prospective uncontrolled real-world studies designed to evaluate quasi-naturalistic product use—are the primary method for assessing the use of premarket tobacco products under ecologically valid conditions. Actual use studies were initially developed for pharmaceuticals to determine whether prescription drugs could be switched to over-the-counter sale by assessing use and safety of the drug in real-world settings designed to mimic over-the-counter use [10,11].

In recent years actual use studies have evaluated use of tobacco products, including oral nicotine pouches [12,13] and heated tobacco (heat-not-burn) products [14,15], among adults who smoke to prospectively assess changes in cigarette smoking following quasi-naturalistic use of the new tobacco products in real-world settings. In contrast to randomized cessation trials, actual use studies do not require participants to have active plans to quit smoking to qualify, do not set explicit goals for product use or smoking, and do not provide instructions, encouragement, or behavioral support to quit. Thus, actual use studies are particularly well-suited to evaluate the ability of ENDS to facilitate switching behavior (smoking abstinence) following periods of unguided self-determined use.

The JUUL2 system is a next-generation pod-based ENDS product that is designed to facilitate high levels of complete switching away from cigarettes among adults who smoke cigarettes. Pharmacokinetic studies show that JUUL2 18 mg/mL (1.5% nicotine by weight) products can deliver similar levels of nicotine as ENDS products with higher nicotine concentrations due to increased aerosol production [16], and behavioral pharmacology data demonstrate that Polar Menthol–flavored 18 mg/mL JUUL2 pods are rated as significantly more satisfying than menthol-flavored pod-based ENDS with 5% nicotine concentrations [17]. The JUUL2 system is currently marketed in the United Kingdom; however, there is a lack of real-world data on the switching potential of these JUUL2 products among US adults who smoke.

Objectives

The primary aim of this actual use study was to evaluate complete switching away from cigarettes (ie, past-7-day and past-30-day smoking abstinence) among US adults who smoked cigarettes every day and were provided with JUUL2 products in their preferred flavor for 6 weeks. Secondary aims of the study were to assess: (1) changes in dependence and respiratory symptoms among participants who switched completely away from smoking at week 6; (2) changes in cigarette consumption among participants who continued smoking at week 6; (3) patterns of JUUL2 product use and subjective responses to JUUL2 products; and (4) factors associated with switching behavior, including JUUL2 study product flavor (menthol vs tobacco), cigarette flavor of usual brand (menthol vs nonmenthol), and baseline ENDS use (dual use along with cigarette smoking).

Methods

Design

The study was designed as an actual use study—a prospective real-world study intended to simulate quasi-naturalistic product use—in which US adults who smoked cigarettes were provided with JUUL2 study products in participant-selected flavors for 6 weeks of ad libitum use in their natural environment. All potential participants first completed an initial prescreening survey (in person or via telephone) and then the formal recruitment screening on the web. The study had 2 distinct arms, implemented at separate research sites with different geographic catchment areas that tested 1 of the 2 following flavor series: (1) traditional flavors (Virginia Tobacco and Polar Menthol), which correspond to ENDS flavors that have been authorized for marketing by US Food and Drug Administration (FDA; 10 sites); or (2) complex flavors (Autumn Tobacco, Summer Menthol, and Ruby Menthol), which represent adult-oriented flavor mixtures that have not yet received market authorization from FDA (14 sites). The 2 arms were conducted at distinct sites and did not share participant pools to evaluate traditional and complex flavors independently. The decision to separate the traditional and complex flavors was an a priori design based on differences in the composition of traditional and complex flavors (see Study Products section), with implied different regulatory positions (ie, FDA has authorized tobacco- and menthol-flavored ENDS, but has not yet authorized nontobacco or nonmenthol flavors or blends), and to avoid the complexity of each participant and site testing all 5 flavors. Study design and procedures for the 2 arms were identical, and the geographic and demographic position of the research sites were roughly parallel, but sites recruited participants from their local catchment area, thus there was no overlap in potential participants across study sites. Importantly, participants were not assigned to sites or arm by the study (randomly or otherwise) but were recruited from each site's distinct geographic catchment area. The study sponsor did not play any role in the recruitment of participants and individuals were primarily

recruited from existing databases maintained by the study sites or in person (not sponsor databases).

Eligible participants first completed a 1-week trial period to determine which flavor they would exclusively use for the 6-week actual use period. During the 6-week actual use period, participants completed weekly web-based survey assessments (Multimedia Appendix 1).

At prescreening, participants were informed that the research study involved using new electronic cigarette products for 7 weeks and that participation in the study did not require stopping smoking cigarettes—the study was not framed or described as a study of behavior change or smoking cessation. During the trial week and 6-week actual use period, no goals or instructions were provided regarding cigarette smoking; participants could smoke ad libitum throughout the study, or reduce or forego smoking, at their discretion. Similarly, no instructions (beyond the user guide included in the product package) were provided regarding use of the JUUL2 study products with respect to amount or persistence of use. The only directions were those that come with marketed products, describing how to use the product and how to charge the device.

Participants

The study sample included healthy non–treatment-seeking English-speaking US adults (aged 22 to 65 years) who smoked cigarettes every day and lived close to one of the 24 shopping-mall–based study sites distributed across 15 states (Multimedia Appendix 2). Eligibility criteria were as follows: (1) smoked ≥100 combustible tobacco cigarettes in their lifetime and smoked cigarettes for ≥12 months before screening; (2) smoked cigarettes every day and smoked an average of ≥5 cigarettes per day; (3) interested in trying and using JUUL ENDS products; (4) access to a computer or laptop, smartphone, or tablet with internet access; (5) provided consent for participation and acknowledged willingness and ability to comply with all study requirements. Interest in stopping smoking was not an eligibility criterion.

Exclusion criteria included the following: (1) use of any nicotine replacement therapy (eg, patch and gum) or prescription smoking cessation medications (eg, varenicline and bupropion) within 30 days before screening; (2) diagnosis of a clinically significant medical (eg, asthma, chronic obstructive pulmonary disease, heart disease, high blood pressure, and cancer) or psychiatric disease; (3) was pregnant, nursing, or intended to become pregnant at any time through the end of study; (4) participated in a research study about tobacco products or ENDS within the past 30 days; (5) was a current or former employee or related to a current or former employee of the tobacco or ENDS industry or any vendor associated with conduct of the study; and (6) operated or lived with family member who operated any home childcare or health care services. There were no eligibility criteria regarding prior use of ENDS; participants could be ENDS naive, former ENDS users, or current dual users. Recruitment aimed to enroll a diverse sample with respect to sociodemographic and tobacco product use characteristics, according to the following soft quotas: (1) ≤60% of male or female participants; (2) at least 35% of participants who identified as non-Hispanic White; (3) ≥50% participants aged <45 years; (4) ≥40% of participants who smoked mentholated cigarettes; and (5) ≥50% of participants who had ever used ENDS. Individuals were primarily recruited from existing databases maintained by the study sites and the study sponsor did not play any role in recruitment of participants.

Ethical Considerations

Participants provided written informed consent. All surveys were completed remotely on the web. Participants were compensated US $20 for completing the baseline assessment, US $15 for completing the product trial assessment, and US $25 for completing each weekly assessment. Additional compensation was provided for completing all study assessments and returning study products, with maximum total compensation of US $450. The Advarra Institutional Review Board approved the study protocol (Pro00068019).

Procedure

Following enrollment, participants completed the baseline survey on the web and visited their study site, where they were provided with a JUUL2 device, charging dock, and 2 pods of each of the flavors in their respective study arm (2 traditional flavors or 3 complex flavors) for a 1-week product trial period. Participants were instructed to use each flavor for at least one day, enough to form an opinion of the product; no other instructions were provided. After the trial period, participants completed a brief web-based survey in which they rated their interest in each of the trialed JUUL2 pod flavors. These responses were used to determine if participants qualified for the 6-week actual use period and which flavor they would use (see Interest in Using and Purchasing JUUL2 Products subsection of Measures).

At the start of week 1, participants received an additional JUUL2 device. Every 2 weeks (at the start of week 1, week 3, and week 5) participants were provided with 20 pods for use during the upcoming 2 weeks (10 pods per week, 60 pods total). Participants were invited by email to complete assessments at the end of each of the study weeks and received compensation upon completion of each weekly assessment. Participants had 3 days from the initial invitation to complete the weekly surveys and they received multiple reminder emails prompting them to complete the assessment (surveys not completed within this window were considered missing).

Participants were instructed in the informed consent form and by study staff to inform study staff of any adverse events (AEs) and intercurrent illnesses experienced during the study. In addition, a specific inquiry regarding AEs (ie, if participants had experienced any changes to their health [yes or no]) was made at the product trial (week 0), week 3, and week 6 surveys; at the week 1, 2, 4, and 5, surveys participants were prompted to report any changes to their health to the study staff. Reported health events were assessed and classified via clinical interviews by trained medical personnel.

Study Products

The JUUL2 system is closed-system pod-based ENDS that is inhalation-actuated and does not have any user-modifiable settings, controls, or buttons. Preclinical studies demonstrate

that the aerosol produced by the JUUL2 system contains substantially lower levels of harmful and potentially harmful constituents than cigarette smoke [18]. The 5 JUUL2 pod flavors evaluated in this study are all currently marketed in the United Kingdom. Virginia Tobacco and Polar Menthol have traditional tobacco and menthol flavors, respectively. The 3 complex flavors (Autumn Tobacco, Summer Menthol, and Ruby Menthol) consist of a tobacco or menthol base, respectively, with top notes of fruit: apple for Autumn Tobacco, tropical fruit for Summer Menthol, and berry for Ruby Menthol.

Measures

Participants Sociodemographic and Tobacco Use Characteristics

Sociodemographic and tobacco or nicotine product use characteristics were assessed in the recruitment screener and baseline assessment with items and measures adapted from national surveys, with some modified for relevance.

Interest in Using and Purchasing JUUL2 Products

At the end of the trial week, participants were asked 2 questions about each of the JUUL2 products they used as follows: (1) "How interested are you in using the JUUL2 study product for the next 6 weeks?" (5-point response scale, from 1 ["not at all interested"] to 5 ["extremely interested"]) and (2) "If the JUUL2 study products you tried this week were available for purchase, would you buy one or more of the products?" (4-point response scale, from 1 ["definitely would not buy"] to 4 ["definitely would buy"]). Only participants who responded 3 or higher to both items for at least one flavor continued to the 6-week actual use period. This was intended to mirror real-world self-selection for product use, as individuals who were not interested in using or purchasing a specific flavor of a consumer product would be unlikely to purchase and adopt it. Participants were provided with the single flavor they expressed the greatest interest in using for the 6-week period; those who expressed exactly equivalent interest in multiple flavors were assigned to the flavor group in their respective study arm that was most in need of enrollees, which helped balance the sample size of flavor groups.

Cigarette Smoking

At each weekly survey, participants were asked if they had smoked cigarettes (even one puff) in the past 7 days (yes or no). At the week 6, survey participants were asked if they smoked in the past 30 days: "no" responses were operationalized as switching. Participants who reported smoking in the past 7 days at any survey between weeks 3 and 6 (ie, within 30 days of week 6) were coded as past-30-day smokers, even if they did not complete the week 6 survey or reported past-30-day abstinence at week 6.

Participants who reported smoking in the past 7 days also reported frequency and intensity of cigarette smoking during that period—daily cigarette consumption was calculated as cigarettes smoked per day in the past 7 days, including nonsmoking days (ie, product of the past-7-day frequency and daily intensity).

Dependence on Cigarettes and JUUL2 Products

Dependence on cigarettes was assessed at baseline and dependence on JUUL2 products was assessed at week 6 among participants who had not smoked in the past 30 days with the Tobacco Dependence Index (range 1-5, with higher scores indicating greater levels of dependence), a measure psychometrically validated in the Population Assessment of Tobacco and Health study for cross-product comparisons [19,20].

Respiratory Symptoms

Frequency of respiratory symptoms was assessed at baseline, when all participants smoked cigarettes every day, and at week 6 among the past-30-day switchers using the Respiratory Symptoms Experience Scale (range 1-5, with higher scores indicating greater frequency of respiratory symptoms), which was developed and validated for adults who smoke cigarettes but may not have clinical respiratory disease [21].

Patterns of JUUL2 Product Use

At each weekly survey, participants who reported using JUUL2 products in the past 7 days reported the number of days they used JUUL2 products and completed 2 items that assessed intensity of use as follows: (1) number of use episodes per day and (2) number of puffs per day.

Subjective Responses to JUUL2 Products

Subjective responses to JUUL2 products were assessed at week 6 with the modified Product Evaluation Scale [22].

Statistical Analysis

The study was prospectively powered for its primary aim, to evaluate rates of the past-30-day switching using the Clopper-Pearson exact method [23] with an estimated expected switch rate of 20% (based on switch rates observed in a prior observational study of US adults who purchased JUUL products), achieving a CI of 5%. Power analyses were not conducted for the secondary analyses.

The rate of past-7-day switching was calculated at each weekly survey within each of the 5 flavor groups; the rate of past-30-day switching was also evaluated at the week 6 Survey. A logistic regression model tested the association of JUUL2 flavor group and past-30-day switching at week 6, with Virginia Tobacco as the reference group; separate models contrasted the 2 combined tobacco flavors (Virginia and Autumn Tobacco) to the 3 combined menthol flavors (polar, summer and Ruby Menthol), and the 3 individual menthol flavors, respectively, to Virginia Tobacco. A repeated-measure logistic regression model included the past-7-day switching averaged across all 6 weeks as the dependent variable and JUUL2 flavor group, menthol cigarette preference (mentholated vs nonmentholated), and time (week) since baseline (coded as continuous, range 1-6) as simultaneous regressors—all 2-way interaction terms and the 3-way JUUL2 flavor × cigarette flavor × time interaction term were also included.

A separate repeated-measure logistic regression model was used to assess the association of baseline ENDS use (past 30 days vs not past 30 days) and the past-7-day switching. In addition,

mixed effects models evaluated linear changes over time in metrics of daily JUUL2 product use across the 6-week period.

Paired *t* tests were used to assess within-person changes in levels of dependence from baseline to week 6 (from cigarettes to JUUL2 products) among participants who reported no past-30-day smoking at week 6; dual users were not included in this analysis as they continued to smoke cigarettes and their dependence on JUUL2 could not be isolated. Changes in respiratory symptoms from baseline to week 6 were similarly assessed among complete 30-day switchers to assess whether switching affected their respiratory symptoms. Among participants who reported past-7-day smoking at week 6, changes in daily cigarette consumption (cigarettes per day on smoking days) from baseline were also tested with paired *t* tests. The proportion of participants reporting reductions of ≥50% was also calculated.

The primary analyses of switching used all observed (nonmissing) data without imputing smoking status to missed observations. This approach is typical for observational studies [24], in contrast to smoking cessation treatment trials, where there is concern that participants might avoid reporting smoking to avoid admitting to "failure" in an agreed behavioral goal [25]. To assess the potential for missingness to introduce bias, analyses were performed to compare participant baseline characteristics between participants who completed all 6 weekly surveys and those who missed any follow-ups. As a sensitivity analysis, the primary switching analyses were repeated while imputing smoking to all missing observations (intent-to-treat).

In addition, among participants who smoked mentholated cigarettes, differences in sociodemographic and tobacco use characteristics were compared between those who selected tobacco-flavored and menthol-flavored JUUL2 products.

Data were analyzed using SAS version 9.4 (SAS Institute Inc) with alpha level set to 0.05; no adjustments were made for multiple testing.

Results

Participant Accrual and Disposition

The enrollment flow in the traditional and complex flavor arms is provided in Multimedia Appendices 3 and 4, respectively. In the traditional flavors arm, 1874 adults were invited to complete the prescreener; 72.68% (1362/1874) completed it and were sent the eligibility screener; and 54.99% (749/1362) subsequently completed the recruitment screener and met all eligibility criteria. Of the eligible participants, 667 (89.1%) provided informed consent and were sent the baseline assessment (Multimedia Appendix 3). Approximately 99% (661/667) of those who started the baseline assessment completed it and were invited to participate in the trial week, 78.8% (521/661) subsequently completed the trial week assessment and were eligible to continue into the 6-week actual use period; 87% (575/661) completed the trial week survey; and 8.2% (54/661) were ineligible to continue into the 6-week actual use period based on expressing insufficient interest in using the study products.

In the complex flavors arm 2625 adults were invited to complete the prescreener; 71.7% (1882/2625) completed it and were sent the eligibility screener and 56.7% (1067/1882) subsequently completed the recruitment screener and met all eligibility criteria. Of the eligible participants, 942 (88.3%) provided informed consent and were sent the baseline assessment (Multimedia Appendix 4). Approximately 98% (927/942) of those who started the baseline assessment completed it and were invited to participate in the trial week, 79.9% (741/927) subsequently completed the trial week assessment and were eligible to continue into the 6-week actual use period; 13.6% (126/927) did not complete the trial week survey; and 6.5% (60/927) were ineligible to continue into the 6-week actual use period because they did not express sufficient interest in using the study products.

At screening, 61 (1.9%) individuals were excluded due to lack of interest in trying JUUL products. At the end of the trial week, the mean ratings of interest in purchasing JUUL2 products ranged from 3.37 to 3.44 on the 4-point response scale; ratings of interest in individual JUUL2 flavors ranged from 3.65 (Virginia Tobacco) to 4.12 (Polar Menthol) on the 5-point response scale (Multimedia Appendix 5). In the traditional flavors arm, 42.2% (242/574) of the participants expressed equivalent interest in both flavors (Virginia Tobacco and Polar Menthol), and in the complex flavors arm, 29.3% (232/792) expressed equivalent interest in all 3 flavors—27.3% (216/792) rated 2 of the 3 flavors equally and 43.4% (344/792) expressed a unique preference for a single flavor. There was no significant difference in the proportion of participants in the traditional (54/661, 8.2%) and complex (60/927, 6.5%) flavors arms that did not continue in the study due to lack of interest in using JUUL2 products at the end of the trial week (χ^2_1=1.67, *P*=.20). In the traditional flavors arm, 20.3% (106/521) of the participants met the threshold for inclusion for only one flavor. In the complex flavors arm, 30.6% (227/741) qualified for 1 or 2 flavors (ie, did not meet threshold for at least one flavor) and 8.8% (65/741) qualified for only one flavor.

The analytic sample consisted of 1160 total participants, roughly evenly divided into the 5 flavor groups: 242 participants in the Virginia Tobacco group, 239 in the Polar Menthol group, 219 in the Autumn Tobacco group, 236 in the Summer Menthol group and 224 in the Ruby Menthol group (Multimedia Appendices 3 and 4).

Sample Characteristics

Characteristics of the sample are reported for all participants (Table 1) and separately within each of the 5 individual flavor groups (Multimedia Appendix 6). Baseline characteristics were generally similar between JUUL2 flavor groups, although larger proportions of participants who selected menthol-flavored JUUL2 products smoked mentholated cigarettes (Multimedia Appendix 6). The participants (mean age 39.42, SD 11.03 years) were majority female (641/1160, 55.26%) and primarily self-identified as non-Hispanic White (667/1160, 57.5%); 19.22% (223/1160) self-identified as non-Hispanic Black; and 16.47% (191/1160) self-identified as of Hispanic ethnicity (Table 1). Over half (632/1160, 54.48%) of them reported annual household income less than US $50,000 and 39.57% (459/1160)

did not complete education beyond high school. On average, the participants reported smoking for 16.20 (SD 9.87) years; currently smoked 14.11 (SD 8.96) cigarettes per day and reported high levels of dependence on cigarettes (Tobacco Dependence Index-Cigarettes, mean 3.65, SD 0.81); 73.3% (850/1159) smoked mentholated cigarettes; and only 0.95% (11/1160) planned to quit smoking in the next 30 days at baseline (832/1160, 71.72% planned to never quit smoking).

Nearly three-quarters (825/1160, 71.12%) of the participants had ever used ENDS and, of those, 58.3% (481/825) had used ENDS in the 30 days preceding baseline (ie, were dual users). Dual users, on average, used ENDS 15.25 (SD 9.82) days out of the past 30 days. Among current ENDS users, the use of menthol-flavored products (183/481, 38%) was more common than use of tobacco-flavored ENDS (50/481, 10.4%). JUUL was selected as the primary ENDS brand by 34.5% (166/481) of those currently using ENDS.

Table 1. Sociodemographic and tobacco use characteristics of the overall sample (N=1160).

Sample characteristics	Values
JUUL2 flavor selection, n (%)	
Virginia Tobacco	242 (20.86)
Polar Menthol	239 (20.6)
Autumn Tobacco	219 (18.88)
Ruby Menthol	224 (19.31)
Summer Menthol	236 (20.34)
Sociodemographic characteristics	
Age (y), mean (SD)	39.42 (11.03)
Sex, n (%)	
Male	517 (44.57)
Female	641 (55.26)
Other	1 (0.09)
Prefer not to answer	1 (0.09)
Race or ethnicity, n (%)	
Hispanic	191 (16.47)
Non-Hispanic Black	223 (19.22)
Non-Hispanic other race[a]	66 (5.69)
Non-Hispanic White	667 (57.5)
Unknown	13 (1.12)
Marital status, n (%)	
Married	369 (31.81)
Living with partner	222 (19.14)
Divorced, separated, or widowed	167 (14.4)
Never married	379 (32.67)
Prefer not to say	23 (1.98)
Annual household income, n (%)	
<US $50,000	632 (54.48)
US $50,000-US $99,999	408 (35.17)
≥US $100,000	120 (10.34)
Highest level of education, n (%)	
High school graduate or less	459 (39.57)
Some college or trade school	422 (36.38)
College graduate or more education	279 (24.05)
Employment status, n (%)	
Full-time	754 (65)
Part-time	145 (12.5)
Other	261 (22.5)
Census region, n (%)	
Northeast	54 (4.66)
Midwest	356 (30.69)
South	606 (52.24)
West	144 (12.41)

Sample characteristics	Values
Cigarette smoking characteristics	
Smoke mentholated cigarettes, n (%)	850 (73.34)
Number of cigarettes smoked per smoking day, mean (SD)	14.11 (8.96)
Duration of smoking (y), mean (SD)	16.20 (9.87)
Age started smoking (y), mean (SD)	19.33 (6.27)
Cigarette dependence[b], mean (SD)	3.65 (0.81)
Plan to quit smoking in next 30 days, n (%)	11 (0.94)
Ever plan to quit smoking, n (%)	328 (28.28)
ENDS[c] use characteristics	
Ever used ENDS, n (%)	825 (71.12)
Age first used ENDS (y), mean (SD)	31.31 (11.4)
Ever used ENDS fairly regularly, n (%)	529 (64.12)
Used ENDS in past 30 days, n (%)	481 (58.3)
Number of days used ENDS in past 30 days, mean (SD)	15.25 (9.82)
Number of times used ENDS per use day, median (IQR)	9 (5-18)
ENDS dependence[b], mean (SD)	3.03 (0.97)
Primary ENDS flavor[d], n (%)	
Tobacco	50 (10.4)
Menthol	183 (38.1)
Mint	55 (11.4)
Fruit	152 (31.6)
Dessert or candy	33 (6.9)
Spice or clove	2 (0.4)
Some other flavor	6 (1.3)
Primary ENDS device type[d], n (%)	
Pod based	168 (34.9)
Disposable	229 (47.6)
Tank	56 (11.6)
Mod	8 (1.7)
Primary ENDS brand[d], n (%)	
JUUL	166 (34.5)
Vuse	63 (13.1)
Blu	51 (10.6)
NJOY	29 (6)
Puff bar	67 (13.9)
Other	105 (21.8)

[a]Non-Hispanic other race includes Asian, Pacific Islander or Native Hawaiian, American Indian or Alaska Native, and “another race not listed.”
[b]Tobacco Dependence Index in Population Assessment of Tobacco and Health adult survey (range 1-5; higher scores indicate greater dependence).
[c]ENDS: electronic nicotine delivery systems. Denominator is 481 past-30-day ENDS users with the exception of ever ENDS use.
[d]Participants selected the single flavor, nicotine concentration, or ENDS device they used most often.

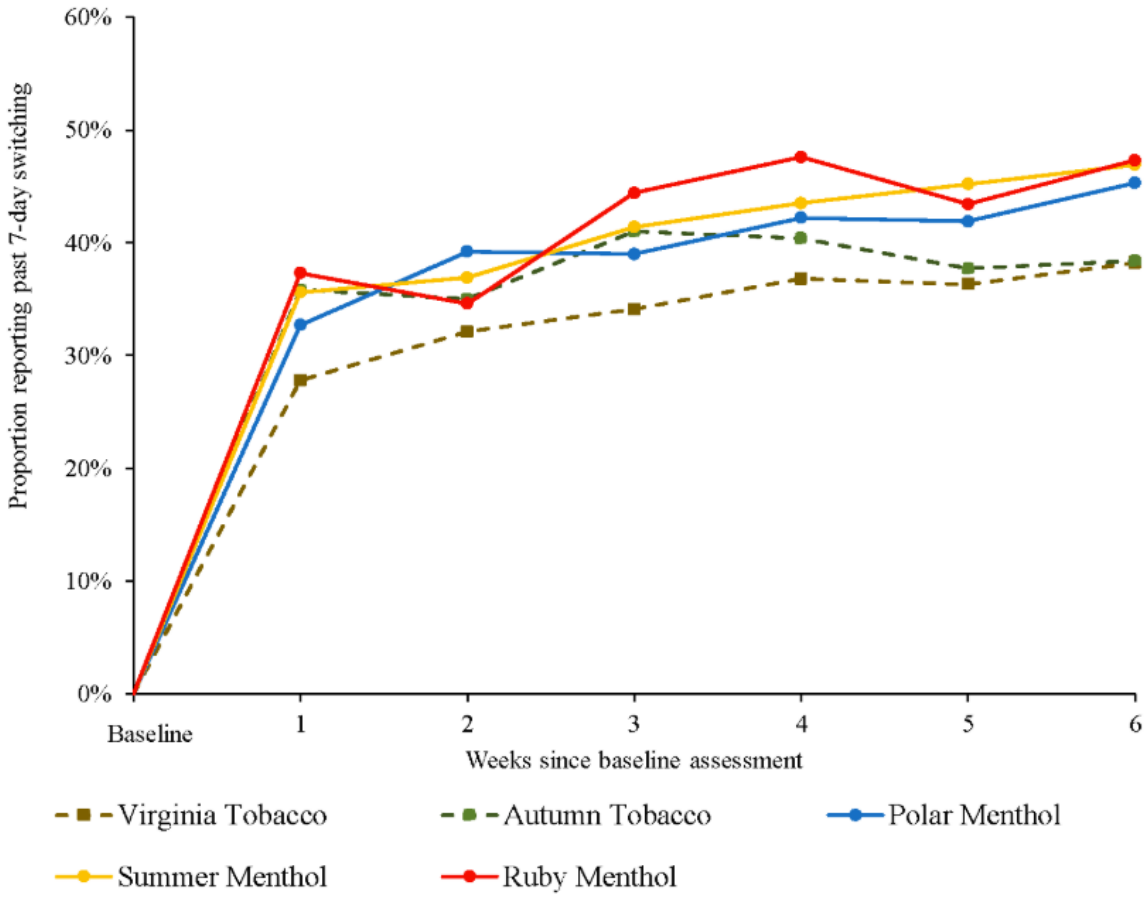
Analysis of Baseline Characteristics by Response to Weekly Surveys

In each flavor group, over 85% of participants who were enrolled and started the 6-week actual use period completed the week 6 survey (Virginia Tobacco, 207/242, 85.5%; Polar Menthol, 212/239, 88.7%; Autumn Tobacco, 190/219, 86.8%; Summer Menthol, 211/236, 89.4%; Ruby Menthol, 201/224, 89.7%) and over 70% completed all 6 weekly surveys (Virginia Tobacco, 178/242, 73.6%; Polar Menthol, 173/239, 72.4%; Autumn Tobacco, 158/219, 72.1%; Summer Menthol, 171/236, 72.5%; Ruby Menthol, 157/224, 70.1%). In the overall sample, there were a few statistically significant differences between the participants who completed all 6 weekly surveys and those who missed some surveys, and the magnitude of differences were small (Multimedia Appendix 7). Participants who completed all 6 surveys (vs 1-5 surveys) were significantly older, by 1.5 years (39.84 vs 38.34 years; *P*=.04), more likely to be female (481/837, 57.5% vs 160/323, 49.5%; *P*=.01), less likely to smoke menthol cigarettes (599/836, 71.7% vs 251/323, 77.7%; *P*=.04) and less likely to have ever used ENDS (575/837, 68.7% vs 250/323, 77.4%; *P*=.003). Among participants who had used ENDS, those who completed all 6 surveys initiated ENDS use about 2 years later (age 31.87 vs 30.01 years; *P*=.03). There were no significant differences by race or ethnicity, education, income, employment, plans to quit smoking, use of JUUL ENDS, or ENDS dependence among those using ENDS. Importantly, there were no differences in the heaviness of smoking or in cigarette dependence, which have been shown to be important predictors of switching. There were no significant differences in rates of survey completion among the JUUL2 flavor groups.

Switching Across 6-Week Actual Use Period

As displayed in Figure 1, in each of the 5 JUUL2 flavor groups, rates of past-7-day switching away from cigarettes increased across the 6-week actual use period: from 27.8% (62/223) at week 1 to 38.2% (79/207) at week 6 for Virginia Tobacco, 35.8% (73/204) at week 1 to 38.4% (73/190) at week 6 for Autumn Tobacco, 32.7% (73/223) at week 1 to 45.3% (96/212) at week 6 for Polar Menthol, 35.6% (77/216) at week 1 to 46.9% (99/211) at week 6 for Summer Menthol, and 37.3% (76/204) at week 1 to 47.3% (95/201) at week 6 for Ruby Menthol. In the combined sample that included all 5 flavors, there was a statistically significant linear increase in the rate of past-7-day switching over the 6-week actual use period (odds ratio [OR] 1.09, 95% CI 1.07-1.12; Multimedia Appendix 8).

Figure 1. Rates of past 7-day switching away from cigarettes across 6-week actual use period by JUUL2 flavor. Virginia Tobacco: week 1: n=223, week 2: n=234, week 3: n=223, week 4: n=220, week 5: n=215, week 6: n=207; Autumn Tobacco: week 1: n=204, week 2: n=206, week 3: n=200, week 4: n=203, week 5: n=191, week 6: n=190; Polar Menthol: week 1: n=223, week 2: n=222, week 3: n=213, week 4: n=218, week 5: n=210, week 6: n=212; Summer Menthol: week 1: n=216, week 2: n=217, week 3: n=210, week 4: n=214, week 5: n=210, week 6: n=211; Ruby Menthol: week 1: n=204, week 2: n=211, week 3: n=205, week 4: n=206, week 5: n=196, week 6: n=201.



At week 6, rates of past-30-day switching were 24.3% (55/226) for Virginia Tobacco, 28.6% (58/203) for Autumn Tobacco, 31.4% (70/223) for Polar Menthol, 33% (69/209) for Ruby Menthol, and 33.9% (74/218) for Summer Menthol (Table 2).

Sensitivity analyses in which smoking (ie, not being switched) was imputed to missing observations showed very similar results, although the absolute imputed switch rates were slightly lower (Multimedia Appendix 9). At week 6, the rates of past-7-day switching using the intent-to-treat approach ranged from 32.6% (79/242; Virginia Tobacco group) to 42.4% (95/224; Ruby Menthol group) and the rates of past-30-day switching ranged from 22.7% (55/242, Virginia Tobacco group) to 31.4% (74/236, Summer Menthol group).

Table 2. Rates of past-30-day switching at week 6 by JUUL2 product flavor.

JUUL2 product flavor	Proportion reporting past-30-day switching, % (SE; 95% CI)
Virginia Tobacco (N=226)	24.3 (2.9; 18.7-29.9)
Autumn Tobacco (N=203)	28.6 (3.2; 22.4-34.8)
Polar Menthol (N=223)	31.4 (3.1; 25.3-37.5)
Summer Menthol (N=218)	33.9[a] (3.2; 27.7-40.2)
Ruby Menthol (N=209)	33[a] (3.3; 26.6-39.4)

[a]Switch rate is significantly higher than Virginia Tobacco group (P<.05).

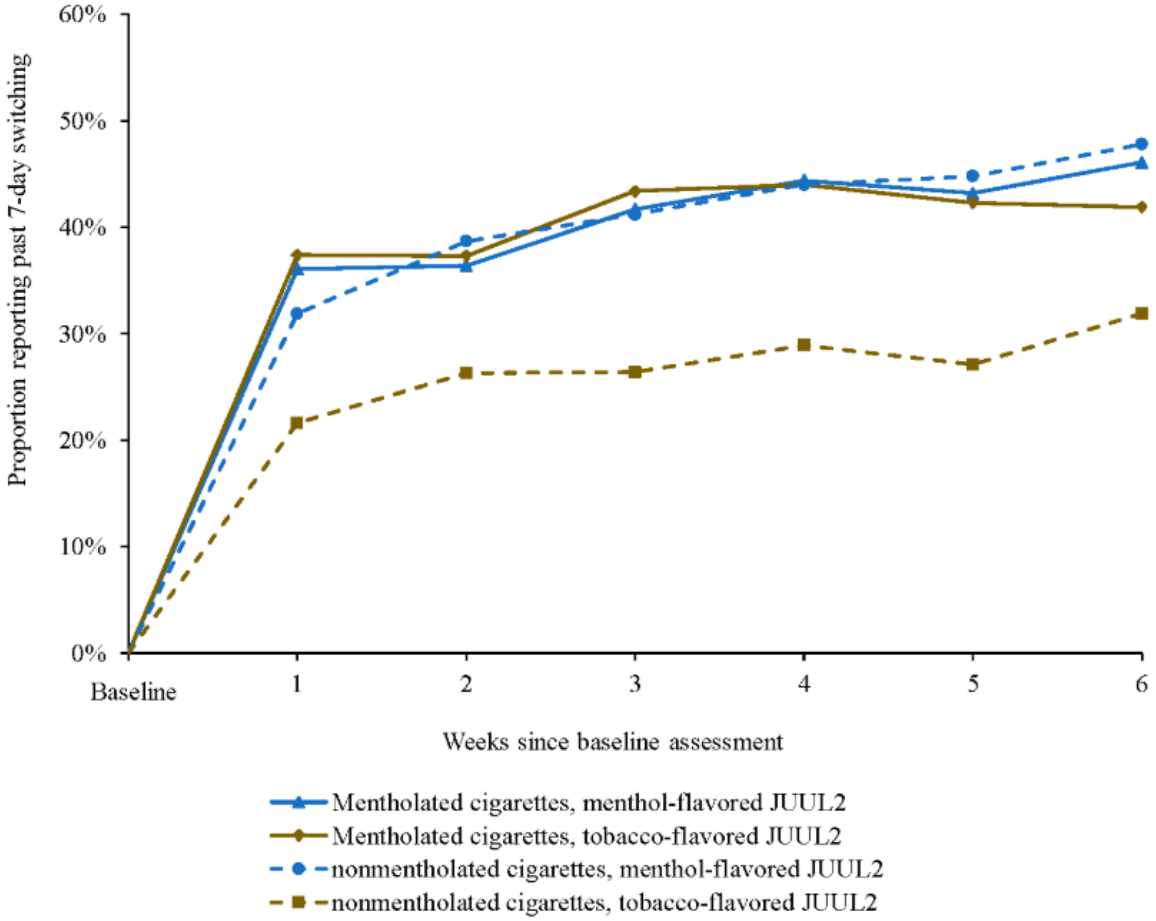
Association of JUUL2 Flavor Group and Menthol Cigarette Smoking With Switching

In the model assessing past-30-day switching at week 6, participants who used menthol-flavored (vs tobacco-flavored) JUUL2 products had statistically significant higher switch rates (OR 1.36, 95% CI 1.04-1.78; Multimedia Appendix 8)—there was no significant main effect of cigarette flavor (P=.06) and the interaction of JUUL2 flavor group and cigarette flavor was not significant (P=.10). Compared to the Virginia Tobacco group, rates of past-30-day switching were significantly higher in the Ruby Menthol group (OR 1.53, 95% CI 1.01-2.33) and Summer Menthol group (OR 1.60, 95% CI 1.06-2.42), respectively. These associations remained significant in the models that imputed smoking for missing data: Ruby Menthol versus Virginia Tobacco (OR 1.51, 95% CI 1.00-2.29) and Summer Menthol versus Virginia Tobacco (OR 1.55, 95% CI 1.03-2.33). The rate of past-30-day switching was numerically higher among the participants in the Polar Menthol group compared to the Virginia Tobacco group, but it did not show a statistically significant difference (OR 1.42, 95% CI 0.94-2.15).

The association of JUUL2 product flavor group and the rate of past-7-day switching was not significant as a main effect (OR 1.23, 95% CI 0.92-1.64; Multimedia Appendix 8). However, this association was significantly moderated by preferred cigarette flavor: among participants who smoked nonmentholated cigarettes, those who used menthol-flavored (vs tobacco-flavored) JUUL2 products had significantly higher switch rates (340/824, 41.3% vs 237/880, 26.9%; OR 1.71, 95% CI 1.04-2.83; Figure 2), whereas there was no significant difference in switch rates by JUUL2 product flavor among participants who smoked menthol cigarettes (1228/2975, 41.3% vs 668/1630, 41%; OR 0.88, 95% CI 0.65-1.18).

There was also a significant main effect of menthol cigarette smoking, with smokers of mentholated (vs nonmentholated) cigarettes demonstrating higher switch rates (OR 1.52, 95% CI 1.14-2.04; Multimedia Appendix 8), largely because of lower switch rates among participants who smoked nonmentholated cigarettes and used tobacco-flavored JUUL2 products (Figure 2). Among participants who smoked mentholated cigarettes, there were no statistically significant differences in sociodemographic or tobacco product use characteristics between those who selected tobacco-flavored and those who selected menthol-flavored JUUL2 products (Multimedia Appendix 10).

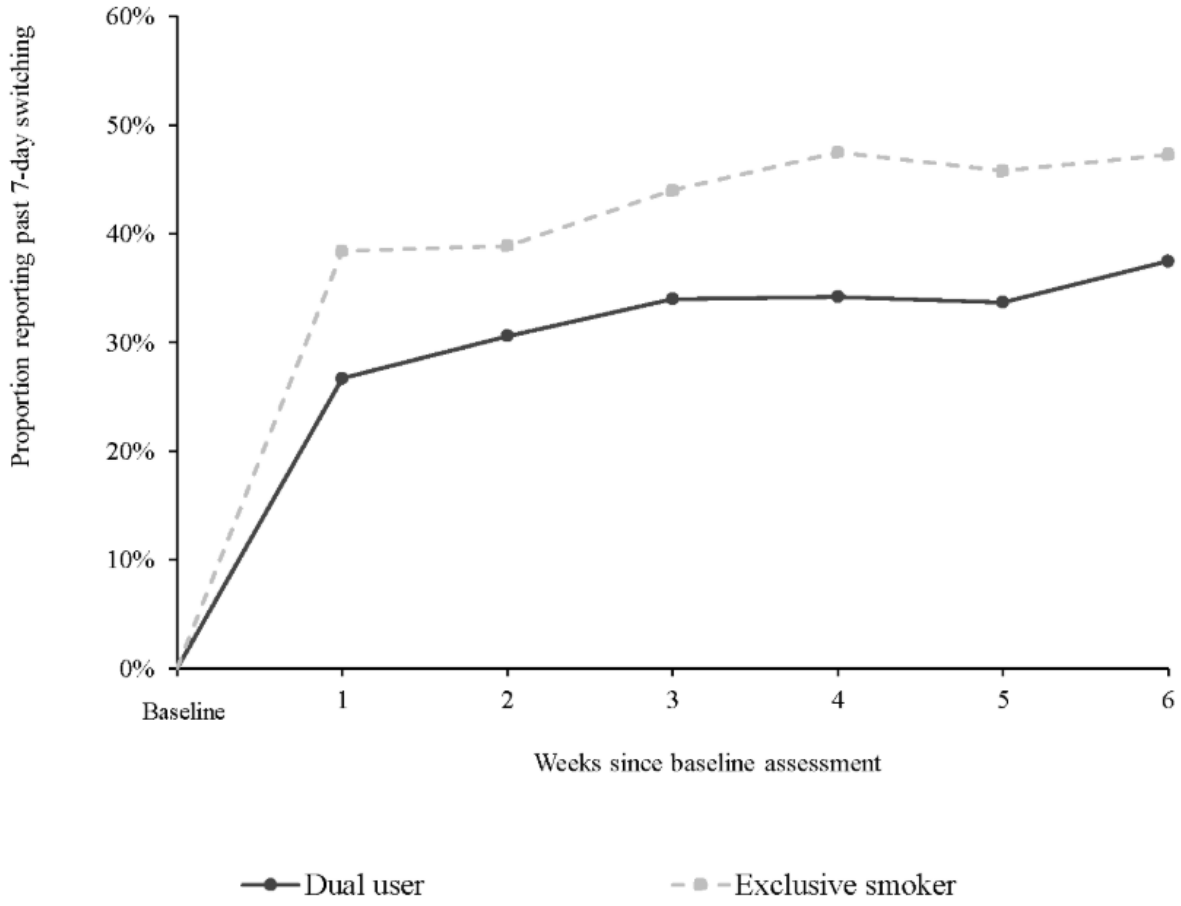
Figure 2. Past 7-day switching by JUUL2 product flavor and cigarette flavor across 6-week actual use period. Smoked menthol cigarettes, menthol-flavored JUUL2: week 1: n=499, week 2: n=508, week 3: n=492, week 4: n=504, week 5: n=482, week 6: n=490; smoked nonmentholated cigarettes, menthol-flavored JUUL2: week 1: n=144, week 2: n=142, week 3: n=136, week 4: n=134, week 5: n=134, week 6: n=134; smoked menthol cigarettes, tobacco-flavored JUUL2: week 1: n=273, week 2: n=287, week 3: n=274, week 4: n=273, week 5: n=265, week 6: N=258; smoked nonmentholated cigarettes, tobacco-flavored JUUL2: week 1: n=153, week 2: n=152, week 3: n=148, week 4: n=149, week 5: n=140, week 6: n=138.



Differences in Switching by ENDS Use at Baseline

There were significant differences in the rates of past-7-day switching according to ENDS use in the 30 days preceding baseline: participants who were smoking but did not use ENDS in the past 30 days had significantly higher switch rates than the dual users (1630/3740, 43.6% vs 843/2575, 32.7%) across the 6-week actual use period (Figure 3 and Multimedia Appendix 11).

Figure 3. Past 7-days switching among baseline dual users and exclusive smokers across 6-week actual use period. Dual users: week 1: n=430, week 2: n=448, week 3: n=435, week 4: n=433, week 5: n=413, week 6: n=416; exclusive smokers: week 1: n=640, week 2: n=642, week 3: n=616, week 4: n=628, week 5: n=609, week 6: n=605.



Changes in Dependence and Respiratory Symptoms Among Participants Who Completely Switched (Did Not Smoke) in the Past 30 Days at Week 6

Among participants who reported past-30-day switching at week 6, levels of dependence on JUUL2 products were significantly lower than their own levels of dependence on combustible cigarettes at baseline, for all 5 flavors (P<.001; Figure 4).

Participants who did not smoke in the past 30 days at week 6 also experienced significant decreases in frequency of self-reported respiratory symptoms relative to baseline when they were smoking cigarettes, across all 5 flavors (P≤.01; Figure 5).

Figure 4. Change from baseline cigarette dependence to JUUL2 dependence at week 6 among the past-30-day switchers (mean, SE). *Statistically significant difference between levels of dependence on cigarettes at baseline and levels of dependence on JUUL2 products at week 6 (*P*<.001).

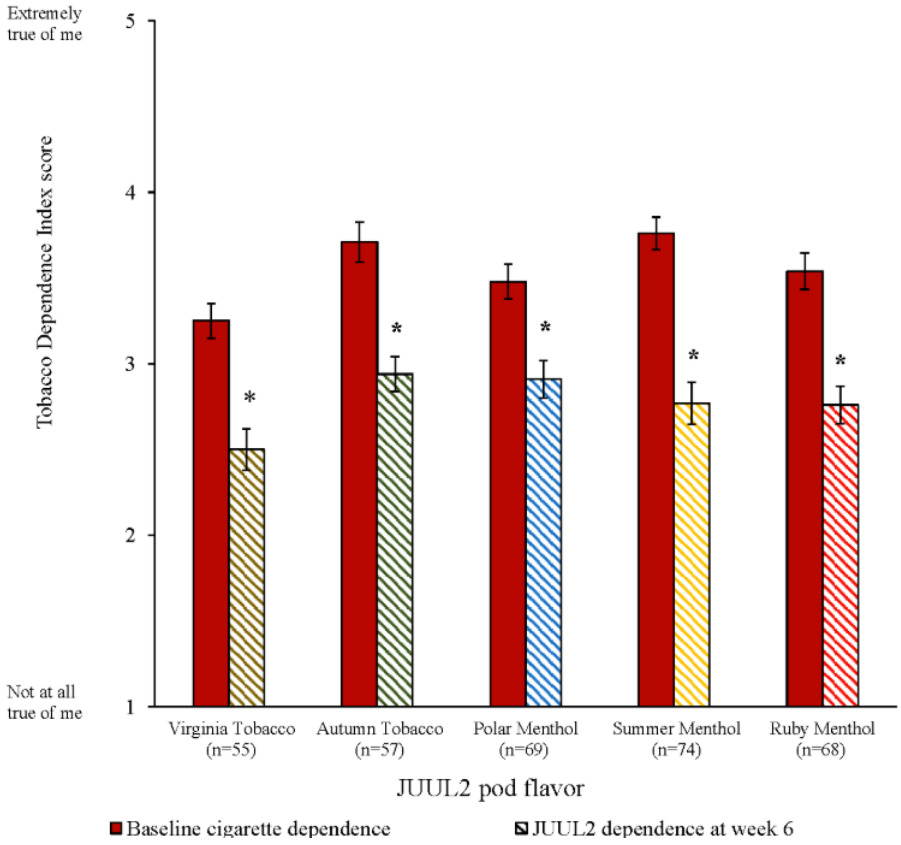
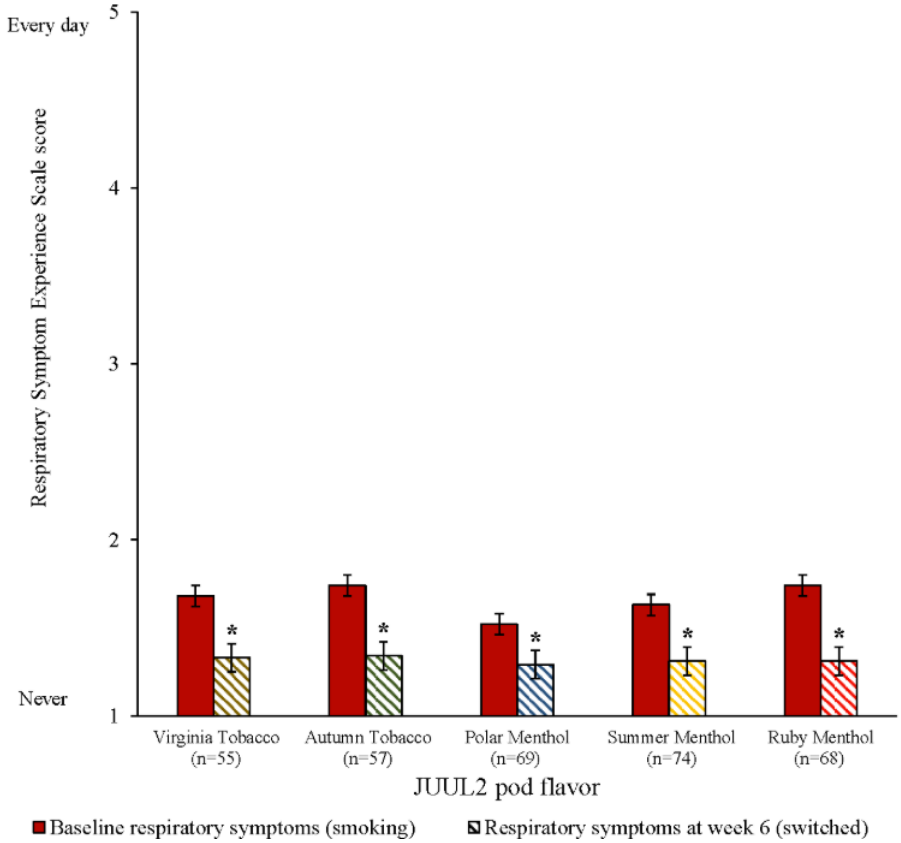


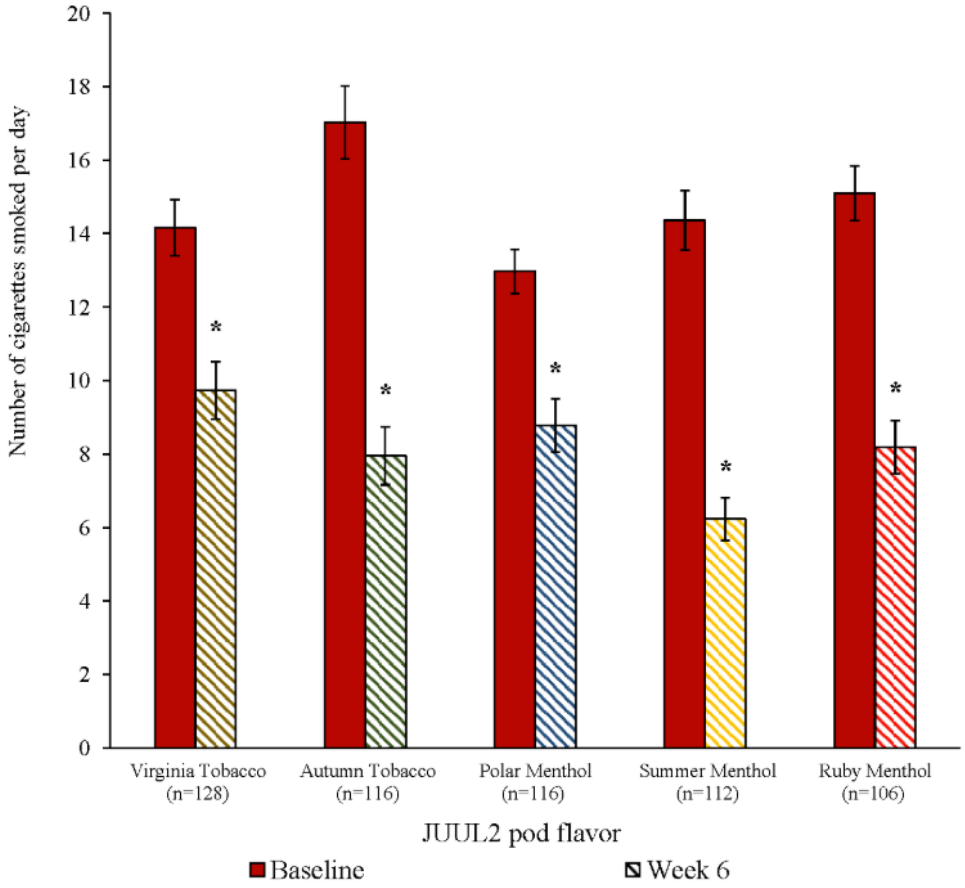
Figure 5. Change in frequency of respiratory symptoms from baseline to week 6 among the past-30-day switchers (mean, SE). *Statistically significant difference between frequency of respiratory symptoms at baseline and at follow-up (P≤.01).



Smoking Reduction Among Participants Who Continued Smoking at Week 6

Among participants who continued to smoke at week 6, average daily cigarette consumption in all 5 JUUL2 flavor groups was significantly reduced relative to participants' own cigarette consumption at baseline (P<.001; Figure 6); across all flavor groups, the proportion of participants that reported reducing their daily cigarette consumption by at least 50% at week 6 relative to baseline ranged from 50.9% (59/116) to 62.9% (73/116; Multimedia Appendix 12).

Figure 6. Change in daily cigarette consumption from baseline to week 6 among participants who did not switch (mean, SE). *Statistically significant difference in number of cigarettes per day smoked at baseline and week 6 among participants who continued smoking (P<.001).

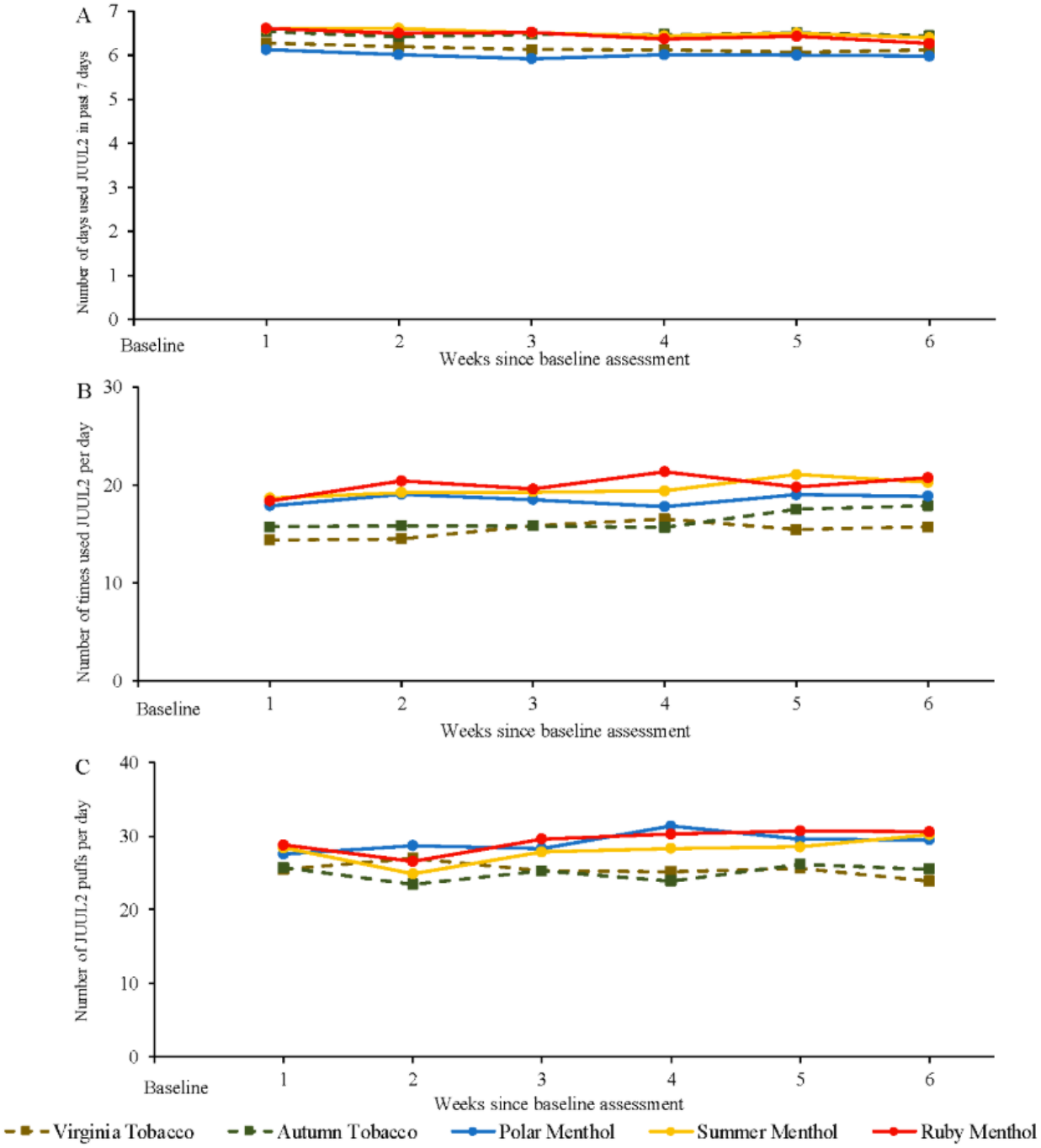


Patterns of JUUL2 Product Use and Subjective Effects

Across the 6-week actual use period, the prevalence of past-7-day JUUL2 product use was over 97% in each of 5 flavor groups (Multimedia Appendix 13). In each of the 5 flavor groups, the median frequency of JUUL2 product use at each weekly survey was 7 days of use in the past 7 days (ie, daily use; Figure 7A); participants generally reported using JUUL2 products between 15 and 20 times per day (Figure 7B) and 25 to 30 puffs per day (Figure 7C). Across the 6-week period, there were no statistically significant linear increases in JUUL2 use occasions per day in the Virginia Tobacco and Polar Menthol JUUL2 groups (P>.15; Multimedia Appendix 14); there were statistically significant linear increases over time in use occasions per day among participants who used Autumn Tobacco, Summer Menthol, and Ruby Menthol JUUL2 products but the magnitude of the increases was small (unstandardized β coefficients=.45−.54; Multimedia Appendix 14). Across the 6-week period, there were no statistically significant increases in JUUL2 puffs per day in the Virginia Tobacco and Autumn Tobacco JUUL2 groups (P>.09); there were statistically significant linear increases over time in puffs per day among participants who used Polar Menthol, Summer Menthol, and Ruby Menthol JUUL2 products but the magnitude of the increases was small (unstandardized β coefficients=.43−.60; Multimedia Appendix 14).

Subjective responses to the 5 JUUL2 products were similar across 5 flavors groups at the week 6 survey: ratings of satisfaction approximated "A lot," ratings of psychological reward approximated "Moderately," ratings of aversion approximated "Very little," and ratings of relief approximated "A lot" (Multimedia Appendix 15). When combined across all 5 flavors, ratings of subjective satisfaction (ie, modified Product Evaluation Scale Satisfaction subscale) at week 2 were significantly associated with higher likelihood of complete past-30-day switching at week 6 (OR 1.25, 95% CI 1.11-1.40).

Figure 7. Patterns of JUUL2 product use across 6-week actual use period. (A) Number of days used JUUL2 in past 7 days; (B) number of times used JUUL2 per day; and (C) number of JUUL2 puffs per day.



Adverse Events

The incidence of study-emergent AEs was low for each of the 5 JUUL2 products during the 6-week actual use period—in each group less than 2.5% of participants reported an AE (Virginia Tobacco, 1/262, 0.4%; for, Autumn Tobacco, 2/237, 0.8%, Polar Menthol, 2/262, 0.8%, Ruby Menthol 2/234, 0.9%, for Summer Menthol, 6/249, 2.4%; Multimedia Appendices 16 and 17). No serious AEs were reported in the study and all evaluable AEs were reported as mild or moderate in intensity and resolved during the study.

Discussion

Principal Findings

This actual use study enrolled a sample of >1000 US adults who smoked cigarettes daily, following them for 6 weeks as they used JUUL2 ENDS with their preferred flavor in their natural environment. The sample was sociodemographically diverse and predominantly included US adults who smoked mentholated cigarettes, were not ready to quit smoking (ie, were not planning to quit within 30 days), and had experience using ENDS. On average, participants were highly dependent on cigarettes when entering the study. The results demonstrate that the use of JUUL2 ENDS products was associated with high rates of

completely switching away from cigarettes among US adults who smoke cigarettes, including both those who smoke mentholated cigarettes and nonmentholated cigarettes and users and nonusers of ENDS. Moreover, participants who reported past-30-day switching at the end of the study experienced meaningful reductions in dependence on tobacco products and improvements in respiratory symptoms during this 6-week period.

Despite a profile that did not favor smoking cessation (eg, high levels of cigarette dependence, almost no participants intending to quit smoking in 30 days, and many never planning to quit), rates of 7-day point prevalence abstinence exceeded 40% after 6 weeks, and at week 6 over 24% of participants in each of the 5 flavor groups (Virginia Tobacco, 55/226, 24.3%; Autumn Tobacco, 58/203, 28.6%, Polar Menthol, 70/223, 31.4%, Ruby Menthol, 69/209, 33%, Summer Menthol, 74/218, 33.9%) reported they had not smoked in the previous 30 days—a criterion often used to evaluate smoking cessation in clinical trials [25]. The proportion reporting switching away from smoking increased over the 6-week period, a pattern that was observed to continue over longer periods in previous observational and randomized studies of JUUL product users [8,26]. Furthermore, the high rates of switching are consistent with participants' reports of high levels of subjective satisfaction in this study and in experimental studies evaluating JUUL2 products [17], which is an established predictor of switching behavior [26] that was confirmed in this study.

It was notable that high switching rates were observed even though participants reported relatively light use of JUUL2, reportedly averaging around 25 to 30 puffs per day. One factor for this observation may be that the participants appear to have spread their use over time, engaging in "grazing" [27]—they reported averaging 15 to 20 use occasions per day, implying just 1 or 2 puffs per occasion. The volume of aerosol produced by the JUUL2 device may also have made puffs more impactful, mitigating a need for more puffs. In any case, more research on patterns of use associated with switching, including using objective measures of puff topography, would be useful.

Participants who used menthol-flavored (vs tobacco-flavored) JUUL2 products had significantly higher rates of past-30-day switching at the end of the 6-week actual use period. In addition, the association of JUUL2 flavor group and past-7-day switching significantly varied by menthol cigarette smoking (ie, the JUUL2 flavor and menthol cigarette interaction term was statistically significant): participants who smoked nonmentholated cigarettes and used menthol-flavored (vs tobacco-flavored) JUUL2 products demonstrated significantly higher switch rates, but this was not true for those who smoked mentholated cigarettes. The finding that participants who smoked nonmentholated cigarettes and used menthol-flavored ENDS had higher switch rates than those using tobacco-flavored ENDS is consistent with observational studies of US adults who smoke and use ENDS, including JUUL products [28,29], and an actual use study that evaluated the use of heated tobacco products [15]. This suggests that the availability of menthol-flavored ENDS may particularly benefit adults who smoke nonmentholated cigarettes.

Participants who smoked cigarettes but did not use ENDS when they entered the study had significantly higher rates of past-7-day switching compared to those who were dual users (ie, used ENDS and smoked cigarettes concurrently) across the 6-week actual use period. Ongoing dual use may be a behavioral marker for difficulty in stopping smoking with ENDS, as the dual users were, on average, relatively frequent ENDS users at baseline but had not yet completely switched. Nonetheless, both baseline dual users and baseline non-ENDS users displayed high rates of past-7-day switching at the end of the study.

Among participants who reported switching for the past 30 days at week 6, levels of dependence on JUUL2 products were statistically significant and meaningfully lower than participants' own levels of dependence on cigarettes when they entered the study; the decrease exceeded the minimally important difference for the measure identified in a prior observational study [30]. This finding concords with a large body of evidence indicating that ENDS, including JUUL products, produce lower levels of dependence than cigarettes [1,30-32]. In addition, consumption of JUUL2 products minimally increased across the 6-week actual use period (increases of less than one use occasion and puff per week). In accordance with data from controlled laboratory studies [17], these results demonstrate that JUUL2 products have lower abuse liability than cigarettes.

Participants who completely switched to JUUL2 products for at least 30 days at the completion of the study also reported significantly lower frequency of respiratory symptoms relative to when they entered the study and were smoking cigarettes. This improvement in respiratory symptoms was statistically significant but modest in magnitude—below the minimally important difference of 0.57 proposed in a prior psychometric validation study (operationalized as the difference in frequency of respiratory symptoms between participants with vs without respiratory symptom-relevant diagnoses, including chronic obstructive pulmonary disease) [21]. However, these improvements in respiratory symptoms occurred after relatively short periods of smoking abstinence and 6 weeks of JUUL2 product use, thus it seems likely that respiratory symptoms will continue to improve over longer periods of switching. For example, a study of adults who switched to JUUL products for an average of 3 years found that they had meaningfully lower respiratory symptoms than a matched comparison group that had continued smoking [33]. In this study, the fact that participants reported low baseline respiratory symptom scores when entering the study (mean <2; approximating ratings of "rarely [1-5 days]"), suggested that changes were limited by floor effects.

Substantial proportions of participants who persisted in smoking (ie, were dual users) after 6 weeks of being provided with JUUL2 products significantly reduced their daily cigarette consumption by at least 50% compared to when they entered the study. These levels of reductions in cigarette consumption have been shown to meaningfully reduce exposure to toxicants in cigarette smoke that are associated with smoking-related diseases [34-36] and have been suggested to reduce risk of smoking-related disease [37]. Finally, the JUUL2 products were well tolerated, producing only infrequent and minor AEs.

Strengths and Limitations

Strengths of the study include the large sociodemographically diverse sample of US adults who smoked every day and reported high levels of cigarette dependence, including large numbers of participants who were not planning to ever quit smoking and smoked mentholated cigarettes. Allowing participants to select their preferred flavor for use during the 6-week actual use period replicated real-world consumer product use and allowed attribution of switching behavior to a single JUUL2 product. However, participants were not allowed to change their selected flavor during the 6-week actual use period, as they would be able to in the real world. In addition, this self-selection of flavors and lack of randomization should be considered when interpreting differences in switch rates between the flavor groups.

Rates of retention were high throughout the 6-week actual use period. In addition, comparisons of sociodemographic and tobacco product characteristics among participants who completed all 6 surveys (vs those who missed some surveys) showed that participants who missed some surveys were not meaningfully different from those who completed all the surveys, and were not heavier or more dependent smokers who would have been expected to have lower rates of switching, suggesting low likelihood of bias due to nonresponse and missing observations. Analyses imputing smoking to all missing observations were performed, and yielded results that paralleled those based on observed data, indicating a lack of material bias in survey response.

Only individuals who reported sufficient levels of interest in using JUUL2 products were eligible to enroll in the study, which may have led to higher switch rates, but also enhanced ecological validity by mirroring real-world use of consumer products, where people buy and adopt products they are interested in, and are not arbitrarily prescribed or assigned products in which they have little interest. Assignment to individual flavors was based on expressed interest, which again is consistent with real-world use. In any case, most participants in both study arms expressed sufficient interest in all the flavors offered to them. Despite this interest in multiple flavors, most of the participants did show substantial interest in one or more flavors offered to them. This is important, as it suggests that making a range of flavors available is likely to increase the proportion of adults who smoke adopt ENDS, thus increasing reach of ENDS and potential switching away from smoking.

Participants were provided with JUUL2 study products at no charge, which is not consistent with the real world, and could have resulted in heavier product use and greater retention or influenced switching. The actual use portion of the study was only 6 weeks in duration and there was no follow-up assessment after completion. Future studies could evaluate trajectories of switching among adults who smoke and adopt JUUL2 products over longer periods. A large study of real-world US adult JUUL purchasers showed that switch rates continued to increase over time, reaching 58% after 2 years of initial purchase [6].

A limitation was the analyses' reliance on self-report; cigarette smoking and JUUL2 product use were not biochemically verified, and respiratory symptoms were evaluated with a self-report measure. Although this is standard practice for actual use studies [12-14] and observational population-based studies more broadly [24], biomarker data could be useful in future studies. In addition, since the primary aim was descriptive (ie, assessing the proportion of smokers who switched), and did not involve any inferential tests, no adjustment was made for multiplicity. Causal inferences regarding switching are limited by the lack of a control group that did not receive JUUL2 products, hence it is not possible to determine how many participants would have switched or stopped smoking if they were not provided with JUUL2 products. However, almost all participants in the sample did not plan to quit smoking in the next 30 days at baseline, and many were not planning to ever quit, suggesting that quit rates would have been very low.

Conclusions

The data from this 6-week actual use study demonstrate that adoption of JUUL2 products is associated with substantial rates of completely switching away from cigarettes among US adults who smoke, and meaningful reductions in cigarette consumption among adults who do not completely switch after 6 weeks. Participants who completely switched from smoking cigarettes to use of JUUL2 products at the end of 6 weeks experienced significant reductions in dependence and improvements in respiratory symptoms. Switch rates were higher among participants who used menthol-flavored (vs tobacco-flavored) JUUL2 products, particularly among adults who smoke nonmentholated cigarettes. Although this study primarily focused on behavioral end points, when considered with data from clinical biomarker studies, these changes in tobacco product use behaviors (ie, complete switching and substantial smoking reduction) are likely to decrease exposure to harmful cigarette-related toxicants, and to ultimately reduce risk of tobacco-related diseases.

Acknowledgments

This study was funded by Juul Labs, Inc. The authors would like to thank Lauren C Smith for her assistance with the conduct and implementation of the study.

Authors' Contributions

NIG led the conceptualization of the manuscript, wrote the manuscript text, and prepared the tables and figures. Authors MAS and SP conducted the analyses and assisted in preparation of the tables and figures. Authors RAB, SS, and AS aided in

conceptualization and writing of the manuscript and provided feedback on drafts. All authors contributed to and approved of the final manuscript.

Conflicts of Interest

NIG and RAB are full-time employees of Juul Labs, Inc. MAS, AS, and SP are full-time employees of Pinney Associates, Inc, and SS is a senior adviser to Pinney Associates, Inc. Pinney Associates has consulted Juul Labs, Inc on tobacco harm reduction since October 2019. As of October 2024, Pinney Associates also consults Philip Morris International solely on US regulatory pathways for noncombustible nontobacco nicotine products. Pinney Associates does not consult on combustible tobacco products. AS also individually provides consulting services to the Center of Excellence for the Acceleration of Harm Reduction through ECLAT Srl, which received funding from the Foundation for a Smoke-Free World.

Multimedia Appendix 1
A schematic diagram of the study design.
[PDF File (Adobe PDF File), 167 KB - ijmr_v14i1e60620_app1.pdf]

Multimedia Appendix 2
The study sites in traditional and complex flavor arms.
[PDF File (Adobe PDF File), 133 KB - ijmr_v14i1e60620_app2.pdf]

Multimedia Appendix 3
Participant flowchart—traditional flavors study arm.
[PDF File (Adobe PDF File), 143 KB - ijmr_v14i1e60620_app3.pdf]

Multimedia Appendix 4
Participant flowchart—complex flavors study arm.
[PDF File (Adobe PDF File), 145 KB - ijmr_v14i1e60620_app4.pdf]

Multimedia Appendix 5
Ratings of interest in JUUL2 products and flavors at the end of trial week.
[PDF File (Adobe PDF File), 102 KB - ijmr_v14i1e60620_app5.pdf]

Multimedia Appendix 6
Sociodemographic and tobacco use characteristics among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 196 KB - ijmr_v14i1e60620_app6.pdf]

Multimedia Appendix 7
Sociodemographic and tobacco use characteristics by number of assessments completed.
[PDF File (Adobe PDF File), 267 KB - ijmr_v14i1e60620_app7.pdf]

Multimedia Appendix 8
Association of JUUL2 flavor group and cigarette flavor with past 7-day and past 30-day switch rates across 6-week actual use period.
[PDF File (Adobe PDF File), 116 KB - ijmr_v14i1e60620_app8.pdf]

Multimedia Appendix 9
Rates of past 30-day and past 7-day switching away from cigarettes across 6-week actual use period by JUUL2 flavor imputing missing as smoking (intent-to-treat).
[PDF File (Adobe PDF File), 122 KB - ijmr_v14i1e60620_app9.pdf]

Multimedia Appendix 10
Sociodemographic and tobacco use characteristics of participants who smoked nonmentholated cigarettes by JUUL2 flavor selection.
[PDF File (Adobe PDF File), 149 KB - ijmr_v14i1e60620_app10.pdf]

Multimedia Appendix 11

Association of past 30-day electronic nicotine delivery systems use at baseline with past 7-day switch rates across 6-week actual use period.
[PDF File (Adobe PDF File), 129 KB - ijmr_v14i1e60620_app11.pdf]

Multimedia Appendix 12
Change in daily cigarette consumption from baseline to week 6 among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 149 KB - ijmr_v14i1e60620_app12.pdf]

Multimedia Appendix 13
Patterns of past 7-day JUUL2 product use across 6-week actual use period among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 178 KB - ijmr_v14i1e60620_app13.pdf]

Multimedia Appendix 14
Temporal trends in daily JUUL2 product use across 6-week actual use period among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 121 KB - ijmr_v14i1e60620_app14.pdf]

Multimedia Appendix 15
Subjective responses to JUUL2 products at week 6 survey among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 134 KB - ijmr_v14i1e60620_app15.pdf]

Multimedia Appendix 16
Adverse events—traditional and complex flavors trial week.
[PDF File (Adobe PDF File), 137 KB - ijmr_v14i1e60620_app16.pdf]

Multimedia Appendix 17
Adverse events over 6-week actual use period among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 179 KB - ijmr_v14i1e60620_app17.pdf]

References

1. National Academies of Sciences, Engineering, and Medicine. Public Health Consequences of E-Cigarettes. Washington, DC: The National Academies Press; 2018.
2. Balfour DJ, Benowitz NL, Colby SM, Hatsukami DK, Lando HA, Leischow SJ, et al. Balancing consideration of the risks and benefits of e-cigarettes. Am J Public Health 2021 Sep;111(9):1661-1672. [doi: 10.2105/AJPH.2021.306416] [Medline: 34410826]
3. Gottlieb S, Zeller M. A nicotine-focused framework for public health. N Engl J Med 2017 Sep 21;377(12):1111-1114. [doi: 10.1056/NEJMp1707409] [Medline: 28813211]
4. Toll BA, Smith TT, King BA. Nicotine e-cigarettes: considerations for healthcare providers. Nat Med 2024 Jun;30(6):1513-1514. [doi: 10.1038/s41591-024-02926-7] [Medline: 38627561]
5. Kasza KA, Tang Z, Seo YS, Benson AF, Creamer MR, Edwards KC, et al. Divergence in cigarette discontinuation rates by use of electronic nicotine delivery systems (ENDS): longitudinal findings from the United States PATH study waves 1-6. Nicotine Tob Res 2025 Jan 22;27(2):236-243. [doi: 10.1093/ntr/ntae027] [Medline: 38566367]
6. Kim S, Goldenson NI, Selya A, Shiffman S. Switching away from smoking and reduction in cigarette consumption among U.S. adult purchasers of the JUUL system across 24 months including diverse subpopulations disproportionately affected by cigarette smoking. Nicotine Tob Res 2024 Aug 22;26(9):1183-1191. [doi: 10.1093/ntr/ntae072] [Medline: 38553983]
7. Pulvers K, Nollen NL, Rice M, Schmid CH, Qu K, Benowitz NL, et al. Effect of pod e-cigarettes vs cigarettes on carcinogen exposure among African American and Latinx smokers: a randomized clinical trial. JAMA Netw Open 2020 Nov 02;3(11):e2026324 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.26324] [Medline: 33206193]
8. Xu Y, Goldenson NI, Prakash S, Augustson EM, Shiffman S. Randomized trial assessing the effect of the JUUL system on switching away from cigarettes and smoking reduction among U.S. adults who smoke cigarettes. Exp Clin Psychopharmacol 2024 Feb;32(1):3-15. [doi: 10.1037/pha0000698] [Medline: 38127516]
9. Carpenter MJ, Wahlquist AE, Dahne J, Gray KM, Cummings KM, Warren G, et al. Effect of unguided e-cigarette provision on uptake, use, and smoking cessation among adults who smoke in the USA: a naturalistic, randomised, controlled clinical trial. EClinicalMedicine 2023 Sep;63:102142 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102142] [Medline: 37753443]
10. Bradford DC, McCamon DL, Page BC, Richardson CA, Hunter KA. The actual use trial: a description of design principles and methods. SelfCare 2010;1(4):117-123 [FREE Full text]

11. Csoke E, Landes S, Francis MJ, Ma L, Teotico Pohlhaus D, Anquez-Traxler C. How can real-world evidence aid decision making during the life cycle of nonprescription medicines? Clin Transl Sci 2022 Jan 15;15(1):43-54 [FREE Full text] [doi: 10.1111/cts.13129] [Medline: 34405554]
12. Campbell C, Feehan M, Kanitscheider C, Makena PS, Cai J, Baxter SA. Designing studies to inform tobacco harm reduction: learnings from an oral nicotine pouch actual use pilot study. JMIR Form Res 2022 Aug 19;6(8):e37573 [FREE Full text] [doi: 10.2196/37573] [Medline: 35984682]
13. Becker E, McCaffrey S, Lewis J, Vansickel A, Larson E, Sarkar M. Characterization of Ad Libitum use behavior of On! Nicotine pouches. Am J Health Behav 2023 Jun 30;47(3):428-449. [doi: 10.5993/AJHB.47.3.1] [Medline: 37596760]
14. Roulet S, Chrea C, Kanitscheider C, Kallischnigg G, Magnani P, Weitkunat R. Potential predictors of adoption of the Tobacco Heating System by U.S. adult smokers: an actual use study. F1000Res 2019 Jul 14;8:214 [FREE Full text] [doi: 10.12688/f1000research.17606.2] [Medline: 31559015]
15. Karelitz JL, He Y, Becker E, Vansickel A. Switching behavior and changes in smoking behavior by menthol cigarette preference and menthol heated tobacco product use among adults in the United States who smoke cigarettes: an actual use study. Harm Reduct J 2025 Feb 20;22(1):19 [FREE Full text] [doi: 10.1186/s12954-025-01170-7] [Medline: 39980000]
16. Goldenson NI, Augustson EM, Chen J, Shiffman S. Pharmacokinetic and subjective assessment of prototype JUUL2 electronic nicotine delivery system in two nicotine concentrations, JUUL system, IQOS, and combustible cigarette. Psychopharmacology (Berl) 2022 Mar 20;239(3):977-988 [FREE Full text] [doi: 10.1007/s00213-022-06100-0] [Medline: 35184228]
17. Goldenson NI, Shiffman S, Sembower MA, Black RA. Assessment of abuse liability and switching potential of menthol-flavored pod-based electronic nicotine delivery systems among US adults who smoke cigarettes. Drug Alcohol Depend 2024 May 01;258:111279 [FREE Full text] [doi: 10.1016/j.drugalcdep.2024.111279] [Medline: 38614020]
18. Cook DK, Lalonde G, Oldham MJ, Wang J, Bates A, Ullah S, et al. A practical framework for novel electronic nicotine delivery system evaluation: chemical and toxicological characterization of JUUL2 aerosol and comparison with reference cigarettes. Toxics 2024 Jan 04;12(1):41 [FREE Full text] [doi: 10.3390/toxics12010041] [Medline: 38250996]
19. Strong DR, Leas E, Noble M, White M, Frissell KC, Glasser A, et al. Predictive validity of the adult tobacco dependence index: findings from waves 1 and 2 of the Population Assessment of Tobacco and Health (PATH) study. Drug Alcohol Depend 2020 Sep 01;214:108134 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108134] [Medline: 32629146]
20. Strong DR, Pearson J, Ehlke S, Kirchner T, Abrams D, Taylor K, et al. Indicators of dependence for different types of tobacco product users: descriptive findings from Wave 1 (2013-2014) of the Population Assessment of Tobacco and Health (PATH) study. Drug Alcohol Depend 2017 Sep 01;178:257-266 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.05.010] [Medline: 28675817]
21. Shiffman S, McCaffrey SA, Hannon MJ, Goldenson NI, Black RA. A new questionnaire to assess respiratory symptoms (the respiratory symptom experience scale): quantitative psychometric assessment and validation study. JMIR Form Res 2023 Apr 14;7:e44036 [FREE Full text] [doi: 10.2196/44036] [Medline: 37058347]
22. Hatsukami DK, Zhang Y, O'Connor RJ, Severson HH. Subjective responses to oral tobacco products: scale validation. Nicotine Tob Res 2013 Jul;15(7):1259-1264 [FREE Full text] [doi: 10.1093/ntr/nts265] [Medline: 23239843]
23. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934 Dec;26(4):404-413. [doi: 10.2307/2331986]
24. Blankers M, Smit ES, van der Pol P, de Vries H, Hoving C, van Laar M. The missing=smoking assumption: a fallacy in internet-based smoking cessation trials? Nicotine Tob Res 2016 Jan;18(1):25-33. [doi: 10.1093/ntr/ntv055] [Medline: 25744969]
25. Piper ME, Bullen C, Krishnan-Sarin S, Rigotti NA, Steinberg ML, Streck JM, et al. Defining and measuring abstinence in clinical trials of smoking cessation interventions: an updated review. Nicotine Tob Res 2020 Jun 12;22(7):1098-1106 [FREE Full text] [doi: 10.1093/ntr/ntz110] [Medline: 31271211]
26. Goldenson NI, Shiffman S, Hatcher C, Lamichhane D, Gaggar A, Le GM, et al. Switching away from cigarettes across 12 months among adult smokers purchasing the JUUL system. Am J Health Behav 2021 May 01;45(3):443-463. [doi: 10.5993/ajhb.45.3.4]
27. Notley C, Ward E, Dawkins L, Holland R. The unique contribution of e-cigarettes for tobacco harm reduction in supporting smoking relapse prevention. Harm Reduct J 2018 Jun 20;15(1):31 [FREE Full text] [doi: 10.1186/s12954-018-0237-7] [Medline: 29921278]
28. Mok Y, Jeon J, Levy DT, Meza R. Associations between e-cigarette use and e-cigarette flavors with cigarette smoking quit attempts and quit success: evidence from a U.S. large, nationally representative 2018-2019 survey. Nicotine Tob Res 2023 Feb 09;25(3):541-552 [FREE Full text] [doi: 10.1093/ntr/ntac241] [Medline: 36250607]
29. Selya A, Kim S, Shiffman S, Goldenson NI. Association of use of menthol- versus tobacco-flavored ENDS with switching completely away from cigarettes and differences by menthol cigarette smoking. Subst Use Misuse 2025;60(3):311-318. [doi: 10.1080/10826084.2024.2422963] [Medline: 39491336]
30. Shiffman S, Goldenson NI. Changes in dependence over one year among US adults who smoke cigarettes and switched completely or partially to use of the JUUL-brand electronic nicotine delivery system. Drug Alcohol Depend Rep 2023 Mar;6:100137 [FREE Full text] [doi: 10.1016/j.dadr.2023.100137] [Medline: 36994368]

31. Shiffman S, Goldenson NI, Hatcher C, Augustson EM. Changes in dependence as smokers switch from cigarettes to JUUL in two nicotine concentrations. Am J Health Behav 2021 May 01;45(3):563-575. [doi: 10.5993/AJHB.45.3.10] [Medline: 33894800]
32. Leavens EL, Nollen NL, Ahluwalia JS, Mayo MS, Rice M, Brett EI, et al. Changes in dependence, withdrawal, and craving among adult smokers who switch to nicotine salt pod-based e-cigarettes. Addiction 2022 Jan;117(1):207-215 [FREE Full text] [doi: 10.1111/add.15597] [Medline: 34105208]
33. Shiffman S, Oliveri DR, Goldenson NI, Liang Q, Black RA, Mishra S. Comparing adult smokers who switched to JUUL versus continuing smokers: biomarkers of exposure and of potential harm and respiratory symptoms. Nicotine Tob Res 2024 Mar 22;26(4):494-502. [doi: 10.1093/ntr/ntad197] [Medline: 37837438]
34. Cohen G, Goldenson NI, Bailey PC, Chan S, Shiffman S. Changes in biomarkers of cigarette smoke exposure after 6 days of switching exclusively or partially to use of the JUUL system with two nicotine concentrations: a randomized controlled confinement study in adult smokers. Nicotine Tob Res 2021 Nov 05;23(12):2153-2161 [FREE Full text] [doi: 10.1093/ntr/ntab134] [Medline: 34161586]
35. Holt NM, Shiffman S, Black RA, Goldenson NI, Sembower MA, Oldham MJ. Comparison of biomarkers of exposure among US adult smokers, users of electronic nicotine delivery systems, dual users and nonusers, 2018-2019. Sci Rep 2023 May 05;13(1):7297 [FREE Full text] [doi: 10.1038/s41598-023-34427-x] [Medline: 37147399]
36. Anic GM, Rostron BL, Hammad HT, van Bemmel DM, Del Valle-Pinero AY, Christensen CH, et al. Changes in biomarkers of tobacco exposure among cigarette smokers transitioning to ENDS use: the population assessment of tobacco and health study, 2013-2015. Int J Environ Res Public Health 2022 Jan 27;19(3):1462 [FREE Full text] [doi: 10.3390/ijerph19031462] [Medline: 35162490]
37. Chang JT, Anic GM, Rostron BL, Tanwar M, Chang CM. Cigarette smoking reduction and health risks: a systematic review and meta-analysis. Nicotine Tob Res 2021 Mar 19;23(4):635-642. [doi: 10.1093/ntr/ntaa156] [Medline: 32803250]

Abbreviations

AE: adverse event
ENDS: electronic nicotine delivery systems
OR: odds ratio

Edited by T de Azevedo Cardoso; submitted 03.07.24; peer-reviewed by T Smith, R Polosa; comments to author 09.12.24; revised version received 28.01.25; accepted 28.02.25; published 26.03.25.

Please cite as:
Goldenson NI, Shiffman S, Sembower MA, Selya A, Pype S, Black RA
Evaluating the Effect of the JUUL2 System With 5 Flavors on Cigarette Smoking and Tobacco Product Use Behaviors Among Adults Who Smoke Cigarettes: 6-Week Actual Use Study
Interact J Med Res 2025;14:e60620
URL: https://www.i-jmr.org/2025/1/e60620
doi:10.2196/60620
PMID:

Associations Between Sleep Duration and Activity of Daily Living Disability Among Older Adults in China: Cross-Sectional Study

Huimin Fan[1], PhD; Weijie Yu[2], BMed; Hongguo Rong[3*], PhD; Xiaokun Geng[1,4,5*], MD, PhD

[1]Department of Neurology, Beijing Luhe Hospital, Capital Medical University, Beijing, China

[2]School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China

[3]Institute for Excellence in Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, No. 11 Beisanhuan East Road, Chaoyang District, Beijing, China

[4]China-America Institute of Neuroscience, Beijing Luhe Hospital, Capital Medical University, Beijing, China

[5]Stroke Center, Department of Neurology, Beijing Luhe Hospital, Capital Medical University, No. 82 Xinhua SouthRoad, Tongzhou District, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xiaokun Geng, MD, PhD
Department of Neurology, Beijing Luhe Hospital, Capital Medical University, Beijing, China

Abstract

Background: China has the largest elderly population globally; the growth rate of the aged tendency of the population was higher than that of Western countries. Given the distinctions in historical, ethnic, and economic status as well as socio-cultural background, Chinese adults had different sleep patterns compared with adults in other countries. Considering the heavy disease burden caused by activities of daily living (ADL) disability, we conducted a cross-sectional analysis using data from the China Health and Retirement Longitudinal Study (CHARLS) to test the hypothesis that individuals with short and longer sleep duration are more likely to have ADL disability.

Objective: ADL disability is a common condition affecting the quality of life among older people. This study aimed to explore the associations between sleep duration and ADL disability among middle-aged and older adults in China.

Methods: This cross-sectional study used data from 17,607 participants from the 2018 CHARLS (from 2018 to 2020), an ongoing representative survey of adults aged 45 years or older and their spouses. Self-reported sleep duration per night was obtained from face-to-face interviews. The ADL was measured using a 6-item summary assessed with an ADL scale that included eating, dressing, getting into or out of bed, bathing, using the toilet, and continence. Multiple generalized linear regression models—adjusted for age, sex, education, marital status, tobacco and alcohol use, depression, place of residence, sensory impairment, self-reported health status, life satisfaction, daytime napping, chronic disease condition, and sample weights—were used.

Results: Data were analyzed from 17,607 participants, of whom 8375 (47.6%) were men. The mean (SD) age was 62.7 (10.0) years. Individuals with 4 hours or less (odds ratio [OR] 1.91, 95% CI 1.60 - 2.27; *P*<.001), 5 hours (OR 1.33, 95% CI 1.09 - 1.62; *P*=.006), 9 hours (OR 1.48, 95% CI 1.13 - 1.93; *P*<.001), and 10 hours or more (OR 1.88, 95% CI 1.47 - 2.14; *P*<.001) of sleep per night had a higher risk of ADL disability than those in the reference group (7 hours per night) after adjusting for several covariates. Restricted cubic splines analysis suggested a U-shaped association between sleep duration and ADL disability. When sleep duration fell below 7 hours, an increased sleep duration was associated with a significantly low risk of ADL disability, which was negatively correlated with sleep duration until it fell below 7 hours (OR 0.83, 95% CI 0.79 - 0.87; *P*<.001). When sleep duration exceeded 7 hours, the risk of ADL disability would increase facing prolonged sleep duration (OR 1.19, 95% CI 1.12 - 1.27; *P*<.001). ADL disability should be monitored in individuals with insufficient (≤4 or 5 hours per night) or excessive (9 or ≥10 hours per night) sleep duration.

Conclusions:: In this study, a U-shaped association between sleep duration and ADL disability was found. Future longitudinal studies are needed to establish temporality and examine the mechanisms of the associations between sleep duration and ADL disability.

(Interact J Med Res 2025;14:e65075) doi:10.2196/65075

KEYWORDS

sleep; sleep duration; activities of daily living; CHARLS; survey; questionnaire; self-reported; gerontology; geriatric; older adult; elder; elderly; aging; ADL; physical function; physical functioning; well-being; association; correlation; China Health and Retirement Longitudinal Study

Introduction

The World Health Organization (WHO) survey of 2022 reported that 46.1% of adults aged 60 years and older are living with a disability globally [1]. Furthermore, according to the WHO projections, there will be 66 million older adults in China who live with a functional handicap by 2050 [1]. Activities of daily living (ADL) disability are one of the most common health problems in older adults, affecting about 1 in 6 older adults worldwide [2]. The prevalence of ADL disability increased with aging and had become a significant factor that increases the risk of mortality in older adults [3,4]. ADL disability seriously affected the quality of life and placed a high burden on care providers and the care system. Current research has focused on the prevention of, rehabilitation from, or management of ADL disability, with growing interest in the role of sleep patterns, such as sleep duration [5-7].

Sleep duration has become a significant public health issue [8], associated with the increasing risk of cardiovascular diseases, stroke, and mortality [9-11]. The WHO launched the ”World Mental Health Report—Transforming Health for All” on June 17, 2022 [12]. This report emphasizes that sleep loss is an extremely important risk factor for mental health issues. At present, the relationship between sleep duration and ADL disability has received much attention [13]. A longitudinal observational study suggested that short sleep duration during inpatient rehabilitation might be against greater functional ability at discharge among individuals with acute stroke [14]. The data from the National Health and Nutrition Examination Survey in the United States found that only short sleep duration was significantly associated with ADL disability [15]. The New Integrated Suburban Seniority Investigation Project study also reported that sleeping <6 hours/day was associated with a higher risk of incident disability among older Japanese people [16]. Although the above study did not find a relationship between long sleep and physical activity, the data from the English Longitudinal Study of Ageing cohort showed that both reduced and increased sleep duration were associated with limited agility and mobility disability [17]. A cross-sectional study in Taiwan also reported long sleep duration had a positive association with grip strength performance among older adults [18]. There is still no unified conclusion in the research on the relationship between sleep duration and ADL disability. So far, research on the relationship between sleep duration and ADL disability is well worth conducting.

China has the largest population of older adults globally; the growth rate of the aged tendency of the population was higher than that of Western countries [19,20]. Given the distinctions in historical, ethnic, and economic status as well as socio-cultural backgrounds, Chinese adults had different sleep patterns compared to those from other countries [21,22]. Considering the heavy disease burden caused by ADL disability, we conducted a cross-sectional analysis using data from the China Health and Retirement Longitudinal Study (CHARLS) to test the hypothesis that individuals with short and longer sleep duration are more likely to have ADL disability.

Methods

Ethical Considerations

The CHARLS protocol was approved by the Peking University Biomedical Ethics Review Board (IRB00001052-11015) [23,24]. This cross-sectional study adhered to the reporting guidelines stipulated by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline [25]. According to the Common Rule (45 CFR part 46), this study was exempt from institutional review board approval and the requirement for informed patient consent because we did not use clinical data or involve human participants.

Study Participants

This cross-sectional study used data from the 2018 survey of the CHARLS conducted from 2018 to 2020, a survey of residents aged 45 years or older in 28 provinces across China [26,27]. The national baseline survey of the CHARLS included 17,708 respondents in 150 counties or districts and 450 villages or urban communities throughout the country. Detailed methods of the CHARLS have been described previously [23]. We applied the latest CHARLS wave in 2018 to 19,816 individuals to investigate the relationship between sleep duration and ADL disability. In the current study, we formulated the inclusion criteria for the study subjects: age ≥45 years, respondents with complete information on sleep duration, the ADL scale, and other covariates. We excluded the following outliers: missing date of sample weights, birth date, sleep duration, and ADL scale.

Sleep Duration

Similar to the English Longitudinal Study of Aging [28] in England, the CHARLS evaluated subjects’ sleep duration by utilizing the questions: “During the past month, how many hours of actual sleep did you get at night (average hours for one night)? (This might be shorter than the number of hours you spend in bed).” The response was an average sleep duration of typical usual weekdays or workdays within a month. Consistent with the previous CHARLS sleep study [29-32], we separated respondents into 7 sleep duration groups (≤4, 5, 6, 7, 8, 9, and ≥10 hours out of each evening) in the research. The CHARLS used this question to assess participants’ daytime napping: “In the past month, how long have you taken on average for a nap after lunch?” On the basis of 4 nap time groups considering the results [33,34], the participants were classified as nonnappers (0 min), short nappers (<30 minutes), moderate nappers (30 - 90 minutes), and extended nappers (>90 minutes).

ADL Disability

The CHARLS employed a 6-item summary derived from an ADL scale to quantify ADL utilizing a comprehensive series of inquiries that asked: "Are you experiencing any difficulties with performing any everyday activity due to physical, mental, emotional, or memory impairments, excluding any that you anticipate will persist for less than three months?" The daily tasks encompassed dressing, bathing or showering, eating, getting into or out of bed, using the toilet, and controlling urination and defecation. The response scale contained 4 options: (1) No, I don't have any difficulty; (2) I have difficulty but can still do it; (3) Yes, I have difficulty and need help; and (4) I cannot do it. The ADL scale was extensively used in previous studies among Chinese older adults and has shown good reliability and validity [35,36]. Individuals classified as independent possessed the capability to perform all 6 activities without any difficulty, whereas those who indicated a need for assistance in any of the aforementioned tasks were considered to have an ADL disability [37,38].

Covariates

The CHARLS collected information on sociodemographic characteristics and health-related factors using a structured questionnaire. Covariates that might confound the associations included age, sex, education (elementary school and below, secondary school, and college and above), marital status (married and others), residence (rural or urban), smoking and drinking status (never vs current), visual impairment, hearing impairment, self-reported general health status (ie, very good, good, fair, poor, and very poor), self-reported life satisfaction (ie, completely satisfied, very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied), daytime napping, depression, chronic disease condition (ie, none, mild, and severe), and sample weights. Respondents were deemed to have visual or hearing impairment if they reported fair or poor vision (for either long-distance or near vision) or hearing difficulties. The respondents' chronic disease status was documented based on the count of self-reported noncommunicable diseases, encompassing hypertension, diabetes, dyslipidemia, heart conditions, stroke, renal diseases, asthma, pulmonary disorders, arthritis, liver ailments, and gastric issues. As for sampling weights, the CHARLS constructed cross-sectional sample weights directly from the sampling probabilities for households and individuals, taking into account death and divorce. In the 2018 survey, the CHARLS provides two sets of cross-sectional household weights, one with corrections for nonresponse and one without. Our study adopted the sets of cross-sectional individual weights with corrections.

Statistical Analysis

In the current study, comparisons of respondents' general characteristics based on ADL disability status were conducted using analysis of variance (ANOVA) for numerical variables and ordinal χ^2 tests for categorical variables. Numerical variables were presented as the mean (SD). Categorical variables were reported as numbers and percentages. Given the dichotomous nature of the ADL disability measure utilized in this study, we adopted multivariable generalized linear models with binomial family and log links to examine the associations between sleep duration and ADL disability; meanwhile, the odds ratios (ORs) and 95% CIs were reported. Restricted cubic splines with four knots at the 5th, 35th, 65th, and 95th centiles were used to examine the association between sleep duration per night and ADL disability. The generalized linear analytic models incorporated a range of covariates, including age, sex, education, marital status, tobacco and alcohol consumption, depression, place of residence, sensory impairment, self-reported health status, life satisfaction, daytime napping, chronic disease status, and sample weights. Multiple imputation with 5 replications and a chained equation approach was adopted to deal with missing data. We also conducted analyses stratified by sex to investigate sex-specific associations. A two-sided $P<.05$ was considered to be statistically significant. All data analyses were performed with Stata version 14.0 (StataCorp).

Results

Sample Characteristics

A schematic flow diagram of the study sample is shown in Figure 1. After filtering, our final study sample included 17,607 people with and without ADL disability.

Figure 1. Flowchart of the study sample from the 2018 China Health and Retirement Longitudinal Study (CHARLS).

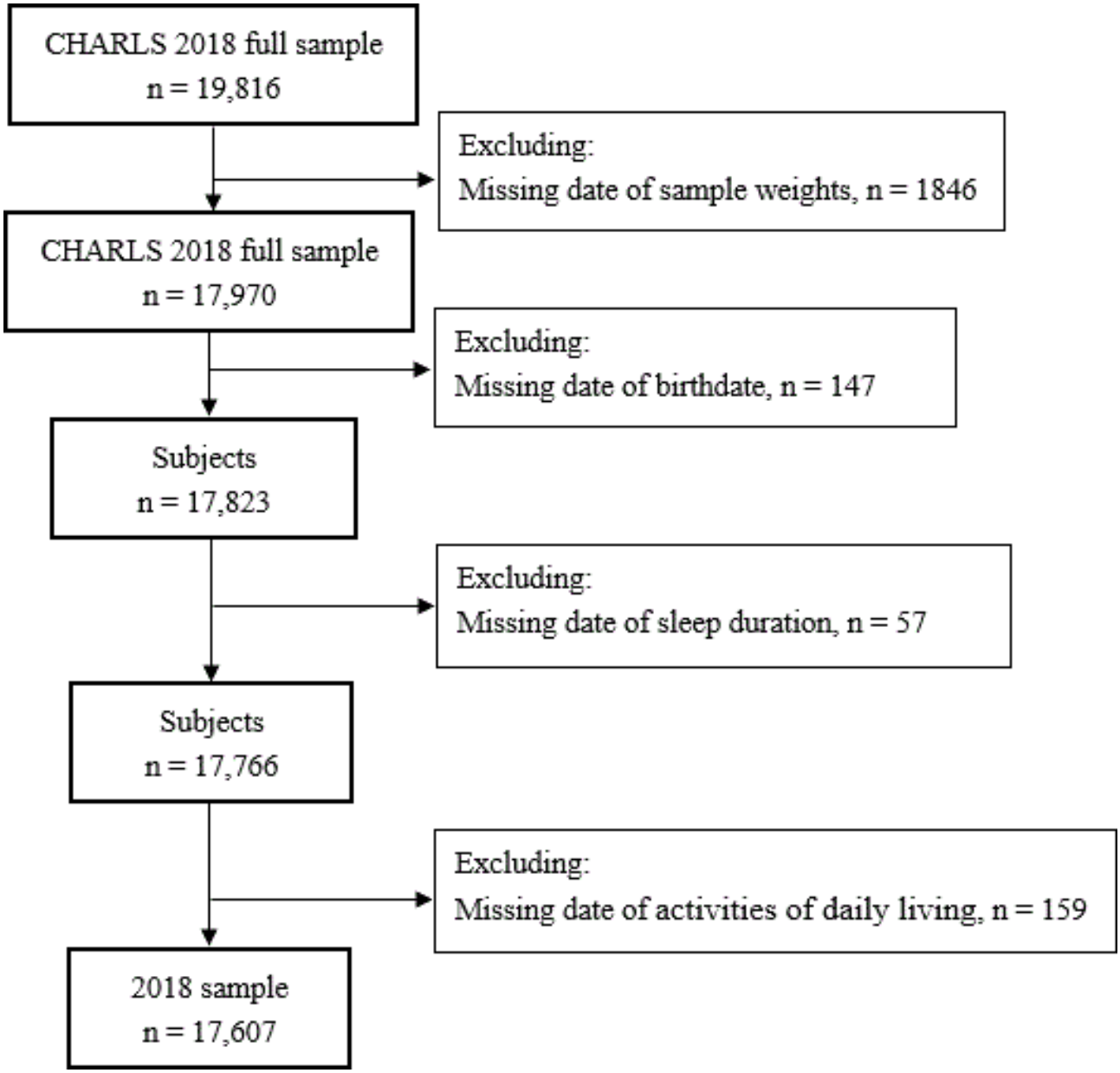


Table 1 shows the characteristics of the study population in the 2018 wave of the CHARLS. Among the 17,607 participants, 8375 (47.6%) were men and 9232 (52.4%) were women, with a mean (SD) age of 62.7 (10.0) years. A total of 3350 participants (19.0%) reported ADL disability. Compared with people with normal ADL function, individuals with ADL disability were more likely to have extremely shorter (≤4 hours) sleep duration and longer (≥10 hours) sleep duration (all P<.001). Participants with ADL disability also tended to have long napping during the day (≥90 minutes). In addition, there were significant differences between ADL disability groups concerning sex, education, tobacco and alcohol use, married status, residence, depressive status, chronic disease conditions, sensory impairments, life satisfaction, and self-reported general health status (all P<.001).

Table . Characteristics of participants by depression status in the 2018 China Health and Retirement Longitudinal Study[a].

Characteristics	Total sample (N=17,607)	ADL[b] disability (n=3350)	No ADL disability (n=14,257)	*P* value
Sleep duration per night, hours, mean (SD)	6.2 (2.0)	5.7 (2.5)	6.3 (1.9)	<.001
Sleep duration per night, hours, n (%)				<.001
≤4	3205 (18.2)	1091 (32.6)	2114 (14.8)	
5	2604 (14.8)	493 (14.7)	2111 (14.8)	
6	3847 (21.9)	552 (16.5)	3295 (23.1)	
7	2936 (16.7)	359 (10.7)	2577 (18.1)	
8	3314 (18.8)	450 (13.4)	2864 (20.1)	
9	804 (4.6)	158 (4.7)	646 (4.5)	
≥10	897 (5.1)	247 (7.4)	650 (4.6)	
Age, mean (SD), y	62.7 (10.0)	68.0 (10.2)	61.5 (9.5)	<.001
Sex, n (%)				<.001
Men	8375 (47.6）	1253 (37.4)	7122 (50.0)	
Women	9232 (52.4）	2097 (62.6)	7135 (50.1)	
Education, n (%)				<.001
Illiterate	4176 (23.7)	1250 (37.3)	2926 (20.5)	
Middle school and below	11,321 (64.3)	1913 (57.1)	9408 (66.0)	
High school and above	2110 (12.0)	187 (5.6)	1923 (13.5)	
Tobacco use, n (%)				<.001
Never	10,074 (57.8)	2063 (62.4)	8011 (56.8)	
Current	7349 (42.2)	1243 (37.6)	6106 (43.3)	
Alcohol use, n (%)				<.001
Never	11,442 (65.0)	2479 (74.0)	8963 (62.9)	
Current	6165 (35.0)	871 (26.0)	5294 (37.1)	
Married, n (%)	13,843 (78.6)	2369 (70.7)	11,474 (80.5)	<.001
Residence, n (%)				<.001
Rural	13,997 (79.7)	2812 (84.0)	11,185 (78.7)	
Urban	3567 (20.3)	534 (16.0)	3033 (21.3)	
Depression, n (%)	6182 (35.1)	1859 (55.5)	4323 (30.3)	
Daytime napping, minutes, n (%)				<.001
None	6727 (38.2)	1301 (38.8)	5426 (38.1)	
≤30	1462 (8.3)	279 (8.3)	1183 (8.3)	
31 - 90	6839 (38.8)	1204 (35.9)	5635 (39.5)	
≥90	2579 (14.7)	566 (16.9)	2013 (14.1)	
Chronic disease conditions, n (%)				<.001
None	9816 (55.8)	1368 (40.8)	8448 (59.3)	
Mild	6736 (38.3)	1595 (47.6)	5141 (36.1)	
Severe	1055 (6.0)	387 (11.6)	668 (4.7)	
Visual impairment, n (%)	14,535 (82.9)	3029 (91.5)	11,506 (80.9)	<.001

Characteristics	Total sample (N=17,607)	ADL[b] disability (n=3350)	No ADL disability (n=14,257)	*P* value
Hearing impairment, n (%)	11,706 (66.6)	2695 (80.7)	9011 (63.3)	<.001
Life satisfaction (satisfied), n (%)	14,365 (81.6)	2267 (67.7)	12,098 (84.9)	<.001
Self-reported general health status (good), n (%)	3846 (21.8)	218 (6.5)	3628 (25.5)	<.001

[a]Missing data for the following characteristics: tobacco use (184, 1.0%), residence (43, 0.2%), visual impairment (77, 0.4%), hearing impairment (35, 0.2%), life satisfaction (1354, 7.7%), and self-reported general health status (1241, 7.0%).

[b]ADL: activities of daily living.

Association Between Sleep Duration and ADL Disability

The association between sleep duration and ADL disability is shown in Table 2. In the model adjusting for age and sex (Model 1), individuals who slept ≤4 hours, 5 hours, 6 hours, 9 hours, and ≥10 hours had a higher risk of ADL disability than the reference group (7 hours per night). After adjusting for age, sex, and other potential confounders (Model 2), the association between 6 hours per night and ADL disability disappeared, while the associations between short (≤4 and 5 hours) and long (9 and ≥10 hours) sleep duration and ADL disability remained the same, with ORs of 1.91 (95% CI 1.60 - 2.27; *P*<.001), 1.33 (95% CI 1.09 - 1.62; *P*=.006), 1.48 (95% CI 1.13 - 1.93; *P*<.001), and 1.88 (95% CI 1.47 - 2.41; *P*<.001), respectively.

Table . Associations between sleep duration and activities of daily living (ADL) disability among participants from the 2018 China Health and Retirement Longitudinal Study.

	OR[a] (95% CI)	*P* value
Sleep duration per night, hours		
Model 1[b]		
≤4	2.85 (2.48 - 3.26)	<.001
5	1.51 (1.30 - 1.76)	<.001
6	1.17 (1.01 - 1.36)	.031
7	[Reference]	-[c]
8	1.08 (0.92 - 1.25)	.342
9	1.41 (1.14 - 1.75)	.002
≥10	2.03 (1.68 - 2.46)	<.001
Model 2[d]		
≤4	1.91 (1.60 - 2.27)	<.001
5	1.33 (1.09 - 1.62)	.006
6	1.11 (0.93 - 1.34)	.246
7	[Reference]	-[c]
8	1.06 (0.88 - 1.27)	.549
9	1.48 (1.13 - 1.93)	<.001
≥10	1.88 (1.47 - 2.41)	<.001

[a]OR: odds ratio.

[b]Model 1 was adjusted for age and sex.

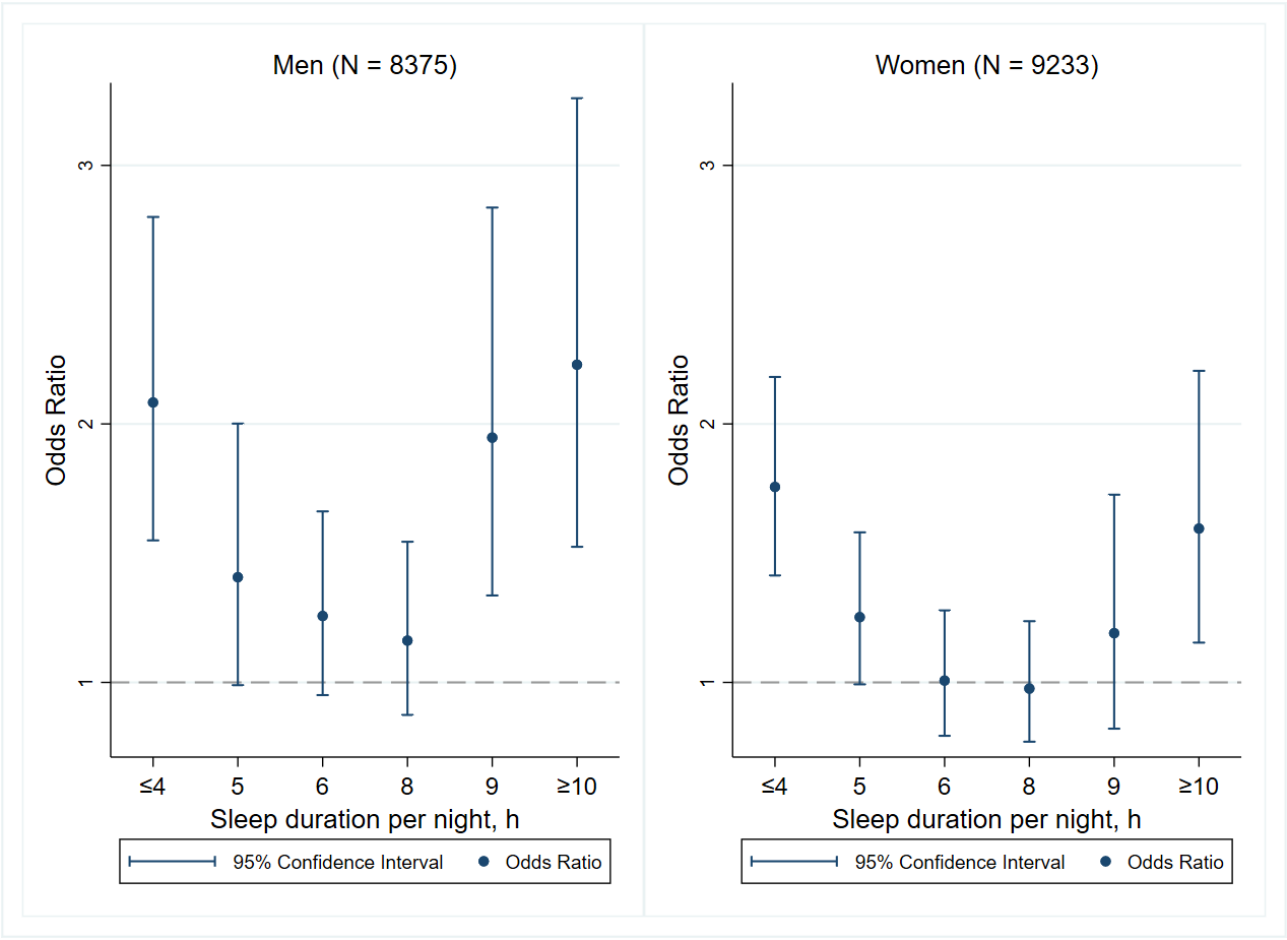
[c]Not applicable.

[d]Model 2 was adjusted for age, sex, education, marital status, tobacco use, alcohol use, afternoon napping, residence, depression, chronic disease conditions, visual impairment, hearing impairment, life satisfaction, self-reported health status, and sample weights.

Sensitivity Analyses

Figure 2 provides the results of stratified analysis by sex. Our findings found that for both men and women, short (≤4 hours) and long (9 and ≥10 hours) sleep duration were associated with higher ADL disability compared to the reference group. However, after adjusting for covariates, men with less than 4 hours (OR 2.08, 95% CI 1.55 - 2.80), 9 hours (OR 1.95, 95% CI13.4 - 2.84), and above 10 hours per night (OR 2.23, 95% CI 1.52 - 3.26) had a higher risk than women did.

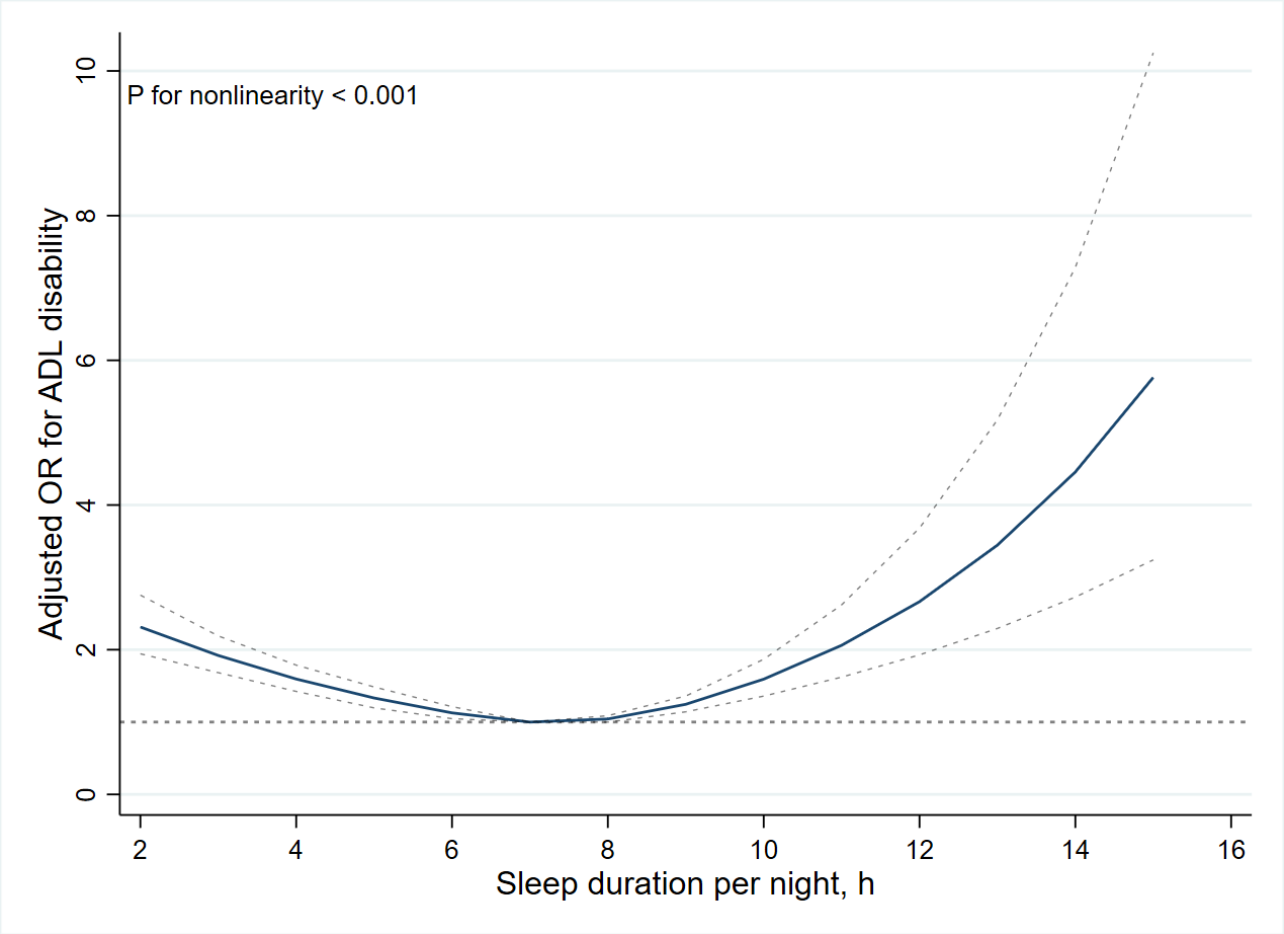
Figure 2. Sex-specific effect of sleep duration on the activity of daily living disability.



Nonlinear Relationship Between Sleep Duration and ADL Disability

In Figure 3, restricted cubic splines were used to flexibly model and visualize the relations between predicted sleep duration and ADL disability. Multivariable-adjusted restricted cubic splines analyses suggested U-shaped associations between sleep duration and ADL disability. The risk of ADL disability was negatively correlated with sleep duration until it bottomed out at 7 hours (OR 0.83, 95% CI 0.79 - 0.87; *P*<.001). However, when the sleep duration was higher than 7 hours, the risk of ADL disability increased significantly (OR 1.19, 95% CI 1.12 - 1.27; *P*<.001).

Figure 3. Nonlinear relationship of sleep duration and activity of daily living disability. OR: odds ratio; ADL: activities of daily living.



Discussion

Principal Findings and Comparison With Previous Works

In this nationally representative study, using the CHARLS database, a statistically significant U-shaped association was found between sleep duration and ADL disability. Individuals with short and long sleep duration were associated with a higher risk of ADL disability. To the best of our knowledge, the current study stands as one of the most extensive examinations of the correlation between sleep duration and ADL disability in the Chinese population.

Compared to other countries, the prevalence rates of ADL disability among older adults were higher in China [39], with one study observing a rate as high as 41% [40]. In contrast, in America, the percentage of older adults aged 65 - 74 years living with functional disability was 25.4% in 2015 [41]. The measurement of ADL disability serves as a crucial tool for monitoring the efficacy of therapeutic interventions across a range of diseases [42]. Given the well-established link between ADL disability and functional ability, it is important to conduct a study on the relationship between sleep duration and ADL disability.

Previous studies have estimated the correlation between sleep duration and ADL disability. The National Health & Aging Trends Study in the United States has demonstrated that insomnia symptoms were correlated with increased likelihood of functional limitations in later life [43]. The Centers for Disease Control and Prevention reported that adults who slept less than 7 hours per night were more prone to difficulties with daily activities than those who slept 7 - 9 hours per night [44]. Data from the New Integrated Suburban Seniority Investigation Project in Japan also suggested a heightened risk of incident disability in older adults associated with sleeping less than 6 hours per night [16]. However, this study found no significant link with long sleep duration, potentially due to its limited sample size (n=256) and subject detection bias. Conversely, data from 486,619 stroke survivors from the National Health Interview Survey revealed that long sleepers (≥9 hours) were more likely to experience problems with instrumental activities of daily living compared to average sleepers [13]. Our results are in accordance with the studies highlighting the increased risk of ADL disability associated with both short and long sleep. Meanwhile, our findings also align with a previous CHARLS study [45], which indicates that sleep durations of short (4 hours or less) and long (9 hours or more) predicted incident ADL disability.

Our results of multivariable-adjusted restricted cubic splines analyses also observed a U-shaped association between sleep duration and ADL disability. Our results lent support to the conclusion that adults aged 18 - 60 years were recommended to sleep at least 7 hours each night to foster optimal health and well-being, as suggested in previous studies [8,46]. Furthermore, “Healthy China 2030” echoes this advice, advocating for a sleep

duration of 7 - 8 hours per day for Chinese adults. Insufficient or excessive sleep duration per night was associated with an increased risk for coronary heart disease, cognitive decline, and functional limitations [30,45,47]. The evidence from China Kadoorie Biobank indicated that 23% of Chinese adults reported getting below 6 hours of sleep duration and 16% reported exceeding 10 hours [48]. The observed nonlinear association underscores the ongoing necessity for public education regarding sleep health and opportunities for healthcare providers to engage patients in discussions about the significance of sleep and to promote sleep health initiatives. Future research to prevent functional limitations may consider intervention strategies focused on sleep hygiene and inform the development of health policies.

The mechanisms by which sleep duration influences ADL disability remain unclear. Several plausible explanations have been proposed to explain the correlation between sleep duration and ADL disability. One possible explanation is tied to the pathology of age-related macular degeneration pathology. An exploratory study involving 277 patients with age-related macular degeneration found significant alterations in the expression levels of immediate early response 3 (IER-3), tissue inhibitor of metalloproteinase-3 (TIMP-3), beta 3-glucosyltransferase (B3GALTL), hepatic lipase (LIPC), and HtrA serine peptidase 1 (HTRA1) among individuals with sleep deprivation or patients with age-related macular degeneration who experienced an increase in sleep duration [49]. Meanwhile, physical activity markedly contributed to exacerbating the severity of age-related macular degeneration. Furthermore, IER-3 and TIMP-3 also had a role in regulating mechanisms in circadian rhythms [49]. The second explanation links the relationship between sleep duration and ADL disability to psychological distress [50]. Similarly, sleep duration also affected the relationship between ADL disability and psychological distress as a mediating pathway, which might also affect ADL disability by causing psychological distress [6]. However, the underlying mechanism governing the association between sleep duration and ADL disability remained uncertain.

In addition, our stratified analysis revealed an elevated risk of ADL disability among men with short and long sleep durations. Sex differences in sleep duration cannot be ignored, as men generally tend to sleep less than women throughout their lifespan [51]. Our findings align with those from the Swedish panel study of living conditions among the oldest old, which showed stronger associations between ADL and instrumental activities of daily living as well as other health indicators among men compared to women between 1992 and 2002 [52]. A longitudinal observation study of 3609 adults aged 65 - 89 years from the Korean National Health and Nutrition Examination Survey also found that men had 2 - 3 times higher odds of developing ADL disability than women [53]. One plausible explanation for this sex difference could be the more severe impact of conditions such as stroke, cancer, and pulmonary disease on older men compared to women, potentially due to men's greater propensity for alcohol and tobacco consumption, thereby increasing the greater risk of ADL disability.

Strengths and Limitations

Our study boasts significant strengths, notably its utilization of a nationally representative sample of Chinese adults aged 45 years and older, rather than relying on clinical samples or occupational cohorts. Furthermore, to our knowledge, it stands as one of the pioneering population-based studies to incorporate both sleep duration and ADL disability. However, several limitations of the current study merit consideration. First, due to its cross-sectional design, we are unable to establish causal relationships between sleep duration and ADL disability based on the observed associations, necessitating a cautious interpretation of the findings. Second, sleep duration was solely assessed through self-reporting, which may introduce potential biases. Third, the CHARLS questionnaire lacks comprehensive sleep-related information, such as objective assessments of the sleep quality, sleep disturbance, and difficulties initiating sleep, which could potentially act as a confounding factor, influencing our estimates. Finally, the binary classification of ADL disability, rather than a continuous scale, might underestimate the associations between sleep duration and ADL disability. Additionally, despite controlling for a range of covariates in the analysis, we cannot discount the influence of unknown factors in this study.

Conclusions

A nationally representative sample in China revealed a statistically significant U-shaped relationship between sleep duration and ADL disability. The U-shaped pattern underscores the importance of monitoring ADL disability in middle-aged and older individuals who exhibit either insufficient or excessive sleep duration. It is imperative for future longitudinal studies to elucidate the causal or bidirectional links between sleep duration and ADL disability, while also delving into the potential mechanisms underlying this intricate interplay.

Acknowledgments

This work was supported by Fundamental Research Funds for the Central Universities (No. 2022-JYB-PY-013). We thank the China Health and Retirement Longitudinal Study team for providing data and simple training in using the database.

Authors' Contributions

HR and XG provided the conception of the research design. HF conducted the data analysis and wrote the first version of the manuscript. WY, XG, and HR provided comments related to the presentation of the findings and critically reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Global report on health equity for persons with disabilities. World Health Organization. URL: https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/global-report-on-health-equity-for-persons-with-disabilities [accessed 2024-01-25]
2. He S, Craig BA, Xu H, et al. Unmet need for ADL assistance is associated with mortality among older adults with mild disability. J Gerontol A Biol Sci Med Sci 2015 Sep;70(9):1128-1132. [doi: 10.1093/gerona/glv028] [Medline: 25834196]
3. van Dalen-Kok AH, Pieper MJC, de Waal MWM, van der Steen JT, Scherder EJA, Achterberg WP. The impact of pain on the course of ADL functioning in patients with dementia. Age Ageing 2021 May 5;50(3):906-913. [doi: 10.1093/ageing/afaa247] [Medline: 33300044]
4. Falk Erhag H, Wetterberg H, Johansson L, Rydén L, Skoog I. Activities of daily living (ADL) and instrumental activities of daily living (IADL) disability in Swedish 85-year-olds born three decades apart-findings from the H70 study. Age Ageing 2021 Nov 10;50(6):2031-2037. [doi: 10.1093/ageing/afab112] [Medline: 34107021]
5. Štefan L, Vrgoč G, Rupčić T, Sporiš G, Sekulić D. Sleep duration and sleep quality are associated with physical activity in elderly people living in nursing homes. Int J Environ Res Public Health 2018 Nov 9;15(11):2512. [doi: 10.3390/ijerph15112512] [Medline: 30423981]
6. Xiao S, Shi L, Xue Y, et al. The relationship between activities of daily living and psychological distress among Chinese older adults: A serial multiple mediation model. J Affect Disord 2022 Mar 1;300:462-468. [doi: 10.1016/j.jad.2021.12.069] [Medline: 34954333]
7. Suzuki R, Meguro M, Meguro K. Sleep disturbance is associated with decreased daily activity and impaired nocturnal reduction of blood pressure in dementia patients. Arch Gerontol Geriatr 2011;53(3):323-327. [doi: 10.1016/j.archger.2010.11.006] [Medline: 21129792]
8. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of healthy sleep duration among adults--United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Feb 19;65(6):137-141. [doi: 10.15585/mmwr.mm6506a1] [Medline: 26890214]
9. Wang C, Bangdiwala SI, Rangarajan S, et al. Association of estimated sleep duration and naps with mortality and cardiovascular events: a study of 116 632 people from 21 countries. Eur Heart J 2019 May 21;40(20):1620-1629. [doi: 10.1093/eurheartj/ehy695] [Medline: 30517670]
10. Åkerstedt T, Ghilotti F, Grotta A, et al. Sleep duration and mortality - Does weekend sleep matter? J Sleep Res 2019 Feb;28(1):e12712. [doi: 10.1111/jsr.12712] [Medline: 29790200]
11. Zhou L, Yu K, Yang L, et al. Sleep duration, midday napping, and sleep quality and incident stroke: The Dongfeng-Tongji cohort. Neurology (ECronicon) 2020 Jan 28;94(4):e345-e356. [doi: 10.1212/WNL.0000000000008739] [Medline: 31827003]
12. World Mental Health Report: Transforming Mental Health for All: World Health Organization; 2022. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2025-02-21]
13. Seixas AA, Chung DP, Richards SL, et al. The impact of short and long sleep duration on instrumental activities of daily living among stroke survivors. Neuropsychiatr Dis Treat 2019;15:177-182. [doi: 10.2147/NDT.S177527] [Medline: 30655670]
14. Williams-Cooke C, Watts E, Bonnett J, Alshehri M, Siengsukon C. Association between sleep duration and functional disability in inpatient stroke rehabilitation: a pilot observational study. Arch Rehabil Res Clin Transl 2021 Sep;3(3):100150. [doi: 10.1016/j.arrct.2021.100150] [Medline: 34589700]
15. Yong LC, Li J, Calvert GM. Sleep-related problems in the US working population: prevalence and association with shiftwork status. Occup Environ Med 2017 Feb;74(2):93-104. [doi: 10.1136/oemed-2016-103638] [Medline: 27609325]
16. Zhang L, Ukawa S, Zhao W, et al. Daily sleep duration and the risk of incident disability among younger elderly Japanese adults in the New Integrated Suburban Seniority Investigation Project: A prospective study using competing event analysis. Geriatr Gerontol Int 2019 Sep;19(9):945-949. [doi: 10.1111/ggi.13732] [Medline: 31332921]
17. Campanini MZ, Mesas AE, Carnicero-Carreño JA, Rodríguez-Artalejo F, Lopez-Garcia E. Duration and quality of sleep and risk of physical function impairment and disability in older adults: results from the ENRICA and ELSA cohorts. Aging Dis 2019 Jun;10(3):557-569. [doi: 10.14336/AD.2018.0611] [Medline: 31165000]
18. Huang WC, Lin CY, Togo F, et al. Association between objectively measured sleep duration and physical function in community-dwelling older adults. J Clin Sleep Med 2021 Mar 1;17(3):515-520. [doi: 10.5664/jcsm.8964] [Medline: 33135630]
19. Sun R, Cao H, Zhu X, Liu JP, Dong E. Current aging research in China. Protein Cell 2015 May;6(5):314-321. [doi: 10.1007/s13238-015-0145-5] [Medline: 25779341]
20. Fang EF, Scheibye-Knudsen M, Jahn HJ, et al. A research agenda for aging in China in the 21st century. Ageing Res Rev 2015 Nov;24(Pt B):197-205. [doi: 10.1016/j.arr.2015.08.003] [Medline: 26304837]

21. Gu D, Sautter J, Pipkin R, Zeng Y. Sociodemographic and health correlates of sleep quality and duration among very old Chinese. Sleep 2010 May;33(5):601-610. [doi: 10.1093/sleep/33.5.601] [Medline: 20469802]
22. Chen X, Wang R, Zee P, et al. Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (MESA). Sleep 2015 Jun 1;38(6):877-888. [doi: 10.5665/sleep.4732] [Medline: 25409106]
23. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb;43(1):61-68. [doi: 10.1093/ije/dys203] [Medline: 23243115]
24. Wu Z, Zhang H, Miao X, et al. High-intensity physical activity is not associated with better cognition in the elder: evidence from the China Health and Retirement Longitudinal Study. Alzheimers Res Ther 2021 Nov 3;13(1):182. [doi: 10.1186/s13195-021-00923-3] [Medline: 34732248]
25. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499. [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
26. Yao Y, Wang K, Xiang H. Association between cognitive function and ambient particulate matters in middle-aged and elderly Chinese adults: Evidence from the China Health and Retirement Longitudinal Study (CHARLS). Sci Total Environ 2022 Jul 1;828:154297. [doi: 10.1016/j.scitotenv.2022.154297] [Medline: 35288137]
27. Lin L, Wang HH, Lu C, Chen W, Guo VY. Adverse childhood experiences and subsequent chronic diseases among middle-aged or older adults in China and associations with demographic and socioeconomic characteristics. JAMA Netw Open 2021 Oct 1;4(10):e2130143. [doi: 10.1001/jamanetworkopen.2021.30143] [Medline: 34694390]
28. Bloomberg M, Brocklebank L, Hamer M, Steptoe A. Joint associations of physical activity and sleep duration with cognitive ageing: longitudinal analysis of an English cohort study. Lancet Healthy Longev 2023 Jul;4(7):e345-e353. [doi: 10.1016/S2666-7568(23)00083-1] [Medline: 37421962]
29. Jing R, Xu T, Rong H, Lai X, Fang H. Longitudinal association between sleep duration and depressive symptoms in Chinese elderly. Nat Sci Sleep 2020;12:737-747. [doi: 10.2147/NSS.S269992] [Medline: 33117009]
30. Ma Y, Liang L, Zheng F, Shi L, Zhong B, Xie W. Association between sleep duration and cognitive decline. JAMA Netw Open 2020 Sep 1;3(9):e2013573. [doi: 10.1001/jamanetworkopen.2020.13573] [Medline: 32955572]
31. Rong H, Wang X, Lai X, Yu W, Fei Y. Associations between sleep duration and sensory impairments among older adults in China. Front Aging Neurosci 2022;14:910231. [doi: 10.3389/fnagi.2022.910231] [Medline: 35754970]
32. Guo J, Fei Y, Li J, Zhang L, Luo Q, Chen G. Gender- and age-specific associations between sleep duration and prevalent hypertension in middle-aged and elderly Chinese: a cross-sectional study from CHARLS 2011-2012. BMJ Open 2016 Sep 6;6(9):e011770. [doi: 10.1136/bmjopen-2016-011770] [Medline: 27601494]
33. Song Y, Liu H, Gu K, Liu Y. Association between nighttime sleep duration, nap time, and mild cognitive impairment in Chinese older adults: a cross-sectional study. BMC Public Health 2024 Sep 2;24(1):2381. [doi: 10.1186/s12889-024-19900-0] [Medline: 39223492]
34. Wang Y, Jiang G, Hou N, et al. Effects and differences of sleep duration on the risk of new-onset chronic disease conditions in middle-aged and elderly populations. Eur J Intern Med 2023 Jan;107:73-80. [doi: 10.1016/j.ejim.2022.11.005]
35. Feng Z, Li Q, Zhou L, Chen Z, Yin W. The relationship between depressive symptoms and activity of daily living disability among the elderly: results from the China Health and Retirement Longitudinal Study (CHARLS). Public Health 2021 Sep;198:75-81. [doi: 10.1016/j.puhe.2021.06.023] [Medline: 34365109]
36. Chan S, Jia S, Chiu H, et al. Subjective health-related quality of life of Chinese older persons with depression in Shanghai and Hong Kong: relationship to clinical factors, level of functioning and social support. Int J Geriatr Psychiatry 2009 Apr;24(4):355-362. [doi: 10.1002/gps.2129] [Medline: 18773498]
37. Connolly D, Garvey J, McKee G. Factors associated with ADL/IADL disability in community dwelling older adults in the Irish longitudinal study on ageing (TILDA). Disabil Rehabil 2017 Apr;39(8):809-816. [doi: 10.3109/09638288.2016.1161848] [Medline: 27045728]
38. Hu Y, Zhou F, Kaminga AC, Yan S, Hu Z. Associations of depressive symptoms and chronic diseases with activities of daily living among middle-aged and older population in China: A population-based cohort study. Front Psychiatry 2022;13:848255. [doi: 10.3389/fpsyt.2022.848255]
39. Feng Q, Zhen Z, Gu D, Wu B, Duncan PW, Purser JL. Trends in ADL and IADL disability in community-dwelling older adults in Shanghai, China, 1998-2008. J Gerontol B Psychol Sci Soc Sci 2013 May;68(3):476-485. [doi: 10.1093/geronb/gbt012] [Medline: 23525547]
40. Zhang Y, Xiong Y, Yu Q, Shen S, Chen L, Lei X. The activity of daily living (ADL) subgroups and health impairment among Chinese elderly: a latent profile analysis. BMC Geriatr 2021 Jan 7;21(1):30. [doi: 10.1186/s12877-020-01986-x] [Medline: 33413125]
41. Wang T, Wu Y, Wang W, Zhang D. Association between coffee consumption and functional disability in older US adults. Br J Nutr 2021 Mar 28;125(6):695-702. [doi: 10.1017/S0007114520003153] [Medline: 32778181]
42. Lim YH, Baek Y, Kang SJ, Kang K, Lee HW. Clinical application of the experimental ADL test for patients with cognitive impairment: pilot study. Sci Rep 2021;11(1):356. [doi: 10.1038/s41598-020-78289-z]

43. Spira AP, Kaufmann CN, Kasper JD, et al. Association between insomnia symptoms and functional status in U.S. older adults. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 1(Suppl 1):S35-S41. [doi: 10.1093/geronb/gbu116] [Medline: 25342821]
44. Centers for Disease Control and Prevention (CDC). Effect of short sleep duration on daily activities--United States, 2005-2008. MMWR Morb Mortal Wkly Rep 2011 Mar 4;60(8):239-242. [Medline: 21368739]
45. Liu M, Du X, Sun Y, Zhou A, Sun S, Wu Y. The mediating role of cognition in the relationship between sleep duration and instrumental activities of daily living disability among middle-aged and older Chinese. Arch Gerontol Geriatr 2021;94:104369. [doi: 10.1016/j.archger.2021.104369] [Medline: 33556636]
46. Grandner MA, Chakravorty S, Perlis ML, Oliver L, Gurubhagavatula I. Habitual sleep duration associated with self-reported and objectively determined cardiometabolic risk factors. Sleep Med 2014 Jan;15(1):42-50. [doi: 10.1016/j.sleep.2013.09.012] [Medline: 24333222]
47. Lao XQ, Liu X, Deng HB, et al. Sleep quality, sleep duration, and the risk of coronary heart disease: a prospective cohort study With 60,586 adults. J Clin Sleep Med 2018 Jan 15;14(1):109-117. [doi: 10.5664/jcsm.6894] [Medline: 29198294]
48. Chen Y, Kartsonaki C, Clarke R, et al. Characteristics and correlates of sleep duration, daytime napping, snoring and insomnia symptoms among 0.5 million Chinese men and women. Sleep Med 2018 Apr;44:67-75. [doi: 10.1016/j.sleep.2017.11.1131]
49. Sharma K, Singh R, Sharma SK, Anand A. Sleeping pattern and activities of daily living modulate protein expression in AMD. PLoS ONE 2021;16(6):e0248523. [doi: 10.1371/journal.pone.0248523]
50. Zhang C, Xiao S, Lin H, et al. The association between sleep quality and psychological distress among older Chinese adults: a moderated mediation model. BMC Geriatr 2022;22(1):35. [doi: 10.1186/s12877-021-02711-y]
51. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep 2021 Feb 12;44(2):zsaa169. [doi: 10.1093/sleep/zsaa169] [Medline: 32886772]
52. Schön P, Parker MG, Kåreholt I, Thorslund M. Gender differences in associations between ADL and other health indicators in 1992 and 2002. Aging Clin Exp Res 2011 Apr;23(2):91-98. [doi: 10.1007/BF03351074] [Medline: 21743288]
53. Kim IH. Age and gender differences in the relation of chronic diseases to activity of daily living (ADL) disability for elderly South Koreans: based on representative data. J Prev Med Public Health 2011 Jan;44(1):32-40. [doi: 10.3961/jpmph.2011.44.1.32] [Medline: 21483221]

Abbreviations

ADL: activities of daily living
B3GALTL: beta 3-glucosyltransferase
CHARLS: China Health and Retirement Longitudinal Study
HTRA1: HtrA serine peptidase 1
IER-3: immediate early response 3
LIPC: hepatic lipase
OR: odds ratio
TIMP-3: tissue inhibitor of metalloproteinase-3
WHO: World Health Organization

Edited by Z Guo; submitted 04.08.24; peer-reviewed by E Pineda, SF Fung, S Ranjbaran; revised version received 19.01.25; accepted 19.01.25; published 05.03.25.

Please cite as:
Fan H, Yu W, Rong H, Geng X
Associations Between Sleep Duration and Activity of Daily Living Disability Among Older Adults in China: Cross-Sectional Study
Interact J Med Res 2025;14:e65075
URL: https://www.i-jmr.org/2025/1/e65075
doi:10.2196/65075

Investigating the Association Between Mean Arterial Pressure on 28-Day Mortality Risk in Patients With Sepsis: Retrospective Cohort Study Based on the MIMIC-IV Database

Qimin Chen[1], BS; Wei Li[1], BS; Ying Wang[1], MD; Xianjun Chen[1], MD; Dehua He[1], BS; Ming Liu[1], BS; Jia Yuan[1], MD; Chuan Xiao[1], MD; Qing Li[1], BS; Lu Chen[2], MD; Feng Shen[1], PhD

[1]Department of Critical Care Medicine, The Affiliated Hospital of Guizhou Medical University, Beijing Road No. 8, Yunyan District, Guiyang, Guizhou Province, China

[2]Clinical Trials Centre, The Affiliated Hospital of Guizhou Medical University, Guiyang, China

Corresponding Author:
Feng Shen, PhD
Department of Critical Care Medicine, The Affiliated Hospital of Guizhou Medical University, Beijing Road No. 8, Yunyan District, Guiyang, Guizhou Province, China

Abstract

Background: Sepsis is a globally recognized health issue that continues to contribute significantly to mortality and morbidity in intensive care units (ICUs). The association between mean arterial pressure (MAP) and prognosis among patients with patients is yet to be demonstrated.

Objective: The aim of this study was to explore the association between MAP and 28-day mortality in ICU patients with sepsis using data from a large, multicenter database.

Methods: This is a retrospective cohort study. We extracted data of 35,010 patients with sepsis from the MIMIC-IV (Medical Information Mart for Intensive Care) database between 2008 and 2019, according to the Sepsis 3.0 diagnostic criteria. The MAP was calculated as the average of the highest and lowest readings within the first 24 hours of ICU admission, and patients were divided into 4 groups based on the mean MAP, using the quadruple classification approach. Other worst-case indications from the first 24 hours of ICU admission, such as vital signs, severity of illness scores, laboratory indicators, and therapies, were also gathered as baseline data. The independent effects of MAP on 28-day mortality were explored using binary logistic regression and a two-piecewise linear model, with MAP as the exposure and 28-day mortality as the outcome variables, respectively. To address the nonlinearity relationship, curve fitting and a threshold effect analysis were performed.

Results: A total of 34,981 patients with sepsis were included in the final analysis, the mean age was 66.67 (SD 16.01) years, and the 28-day mortality rate was 16.27% (5691/34,981). The generalized additive model and smoothed curve fitting found a U-shaped relationship between MAP and 28-day mortality in these patients. The recursive algorithm determined the low and high inflection points as 70 mm and 82 mm Hg, respectively. Our data demonstrated that MAP was negatively associated with 28-day mortality in the range of 34.05 mm Hg-69.34 mm Hg (odds ratio [OR] 0.93, 95% CI 0.92-0.94; P<.001); however, once the MAP exceeded 82 mm Hg, a positive association existed between MAP and 28-day mortality of patients with sepsis (OR 1.01; 95% CI 1.01-1.02, P=.002).

Conclusions: There is a U-shaped association between MAP and the probability of 28-day mortality in patients with sepsis. Both the lower and higher MAP were related with a higher risk of mortality in patients with sepsis. These patients have a decreased risk of mortality when their MAP remains between 70 and 82 mm Hg.

(Interact J Med Res 2025;14:e63291) doi:10.2196/63291

KEYWORDS

mean arterial pressure; 28-day mortality; sepsis; MIMIC-Ⅳ; retrospective study; Medical Information Mart for Intensive Care IV

Introduction

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection [1-3]. While mortality from sepsis has decreased over time after age standardization, it remains high [2-6]. According to the Global Burden of Disease Study, 48.9 million cases of sepsis were reported worldwide in 2017, accounting for 11 million deaths [7]. Additionally, sepsis is the most expensive disease to treat in the United States, costing $23.7 billion annually and accounting for 6.2 percent of all hospital admissions [8]. In low-income nations with a

high sepsis burden, this cost could be significantly greater [9,10]. Sepsis has thus emerged as a serious public health concern with substantial global implications and economic cost.

Despite significant attempts to provide novel organ support strategies and identify the underlying etiology of sepsis, the mortality rate remains high [5,9]. Research indicates that early treatment of sepsis can improve prognosis [1,11]. Therefore, early detection of possible risk factors is crucial for patients with a poor prognosis [12]. Mean arterial pressure (MAP) is one of the most commonly used parameters for the evaluation of sepsis severity [13]. MAP is the driving pressure of tissue perfusion and plays a key role in maintaining the perfusion of tissue and organs. While autoregulation of regional perfusion may protect vital organs such as the brain or kidney from systemic hypotension, tissue perfusion becomes linearly dependent on arterial pressure below a certain MAP (approximately 60 mm Hg) [13]. However, the optimal range of MAP during sepsis resuscitation remains undetermined. The 2016 and 2021 SSC Surviving Sepsis Campaign guidelines recommend that MAP be maintained at ≥65 mm Hg in septic shock [14,15] for improving tissue perfusion and prognosis [14,15]. However, a study by Vincent et al [16] found that MAP did not have a linear relationship with ICU mortality, and several other studies have shown that increased MAP does not necessarily improve clinical outcomes in patients with sepsis or septic shock [13,17-22]. On the contrary, excessively high blood pressure may increase the incidence of adverse events [17].

Given that the appropriate blood pressure range is yet to be determined in patients with sepsis, and considering limitations in sample sizes, differences in study design, variability in covariate adjustments, and the heterogeneity of patient populations included in previous studies, it is necessary to study and identify the appropriate blood pressure range for patients with sepsis based on large databases. Therefore, our study aims to investigate the relationship between MAP and the 28-day risk of mortality in patients with sepsis using Medical Information Mart for Intensive Care (MIMIC-IV), a large sample of US sepsis databases, to identify the inflection point value associated with a lower 28-day risk of death in sepsis. This large sample size will provide more stable and reliable results, allowing us to gain a better understanding of the relationship between MAP and the 28-day risk of mortality in sepsis.

Methods

Study Design and Setting

This retrospective cohort study analyzed data from the MIMIC-IV database, including patients admitted to Beth Israel Deaconess Medical Center between 2008 and 2019.Our study complied with the RECORD (Research Reports Using Observational Routine Collection of Health Data) statement [23].

Participants

The study population comprised adult patients with sepsis identified through *ICD-9* (*99,591 - 99,592*) and *ICD-10* (*R652*, *R6520*, *R6521*) diagnostic codes. Inclusion criteria were patients aged ≥18 years and with a confirmed sepsis diagnosis. Patients lacking MAP data were excluded. Of the initial 377,207 records, 34,981 patients met the eligibility criteria and were included in the final analysis.

Data Collection and Variables

Data extraction was performed by certified researchers following standardized MIMIC-IV database procedures. The primary exposure variable, MAP, was calculated as the average of highest and lowest readings within 24 hours of ICU admission. While MAP values naturally fluctuate during the course of sepsis, we chose this approach to provide clinically actionable targets, acknowledging that a more complex time-varying analysis might capture additional nuances in the relationship between MAP and outcomes. MAP was analyzed both as a continuous variable and categorized into quartiles (Q1: 34.05 - 69.34 mm Hg; Q2: 69.34 - 74.94 mm Hg; Q3: 74.94 - 81.87 mm Hg; and Q4: 81.87 - 159.47 mm Hg). The primary outcome was 28-day all-cause mortality, recorded as a binary variable (1=death, 0=survival).

Covariates were selected based on clinical relevance and previous literature [3,24-26]: demographic characteristics (gender, age, and race), vital signs (heart rate, respiratory rate, and temperature), laboratory measurements (eg, lactate), disease severity indices such as Charlson Comorbidity Index, sequential organ failure assessment (SOFA) score, and acute physiology and chronic health evaluation (APACHE III score), therapeutic interventions (ie, mechanical ventilation, renal replacement therapy, corticosteroids, vasoactive drugs, immunoglobulin, and antibiotics).

Statistical Analysis

The analytical approach comprised three sequential steps. First, we constructed univariate and multivariate binary logistic regression models with progressive covariate adjustment: model 1 (unadjusted), model 2 (adjusted for demographics), and model 3 (fully adjusted for all covariates). Continuous variables were presented as mean (SD) or median (range), and categorical variables as frequencies (percentages). Second, we employed generalized additive models (GAM) with smooth curve fitting to evaluate potential nonlinear relationships between MAP and mortality. For identified nonlinear associations, we determined inflection points using recursive algorithms and constructed piecewise linear regression models. Third, we conducted sensitivity analyses using alternative outcome measures (30-day mortality), different MAP calculation methods (median values), and varying covariate selection approaches. All analyses were performed using R software (version 4.3.2; R Foundation for Statistical Computing) and Empower Stats software (X&Y Solutions, Inc. Boston, MA) [27], with statistical significance set at $P<.05$ (two-sided).

Missing Data Handling

Patients with missing values for any of the following key variables were excluded from the analysis: MAP measurements, age, SOFA score, lactate levels, and mortality outcomes. To assess potential selection bias, we compared the characteristics of included and excluded patients.

Ethical Considerations

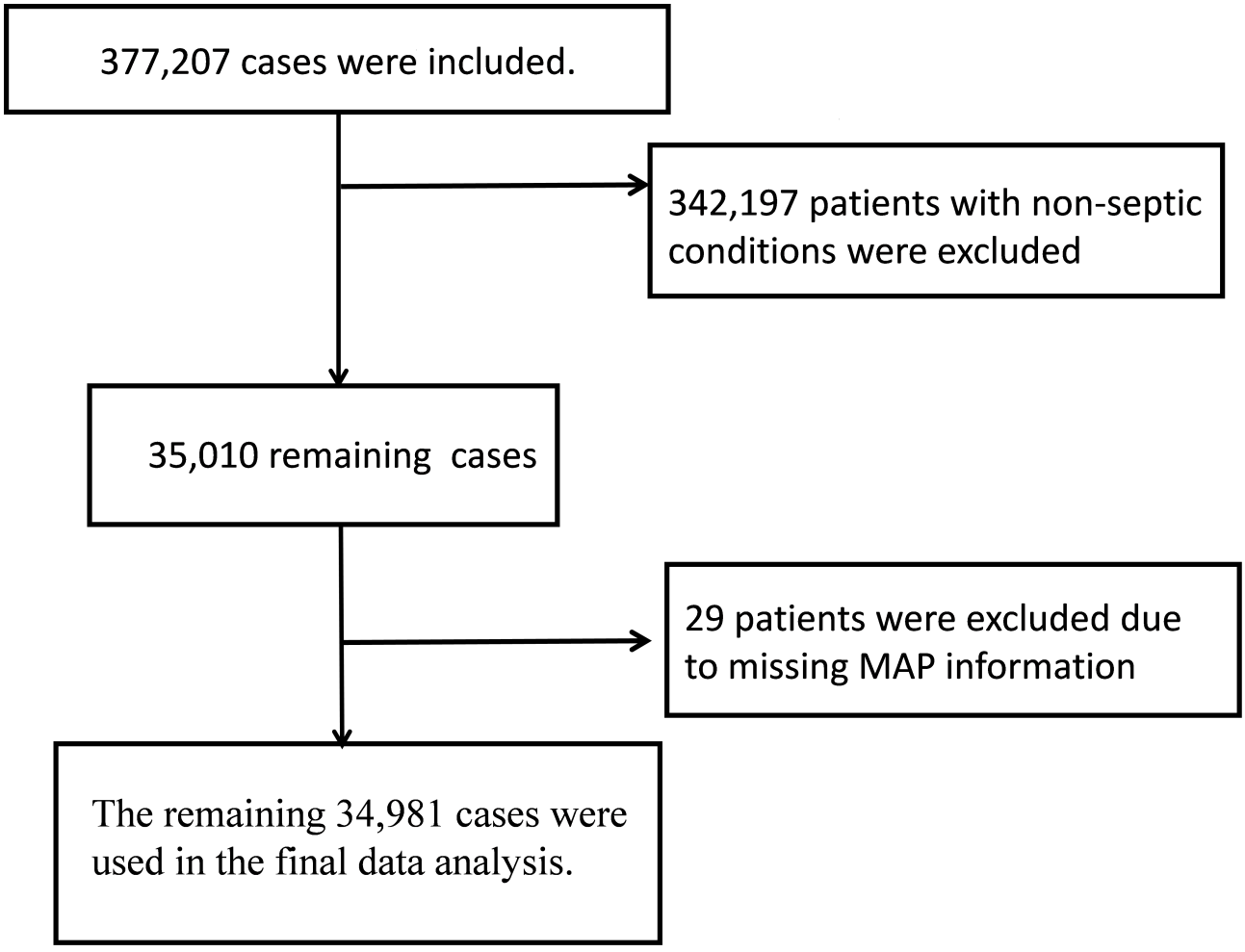
The study protocol received approval from the institutional review boards of Beth Israel Deaconess Medical Center (2001-P-001699/14) and Massachusetts Institute of Technology (0403000206). The data used in this study were obtained from the MIMIC-IV database, which is a freely accessible, deidentified public database [28]. Given the deidentified nature of the database, the requirement for informed consent was waived. Participants were not financially compensated, as the study used data from publicly accessible databases rather than data provided directly by participants.

Results

Study Population Screening

This study initially comprised 377,207 patients, of which 342,197 patients without sepsis and 29 patients lacking MAP data were excluded, with 34,981 cases remaining for final data analysis. The patient selection process is shown in the flowchart (Figure 1). Missing data were minimal (<5%, range 0 - 4.1%) across all variables. Given the low proportion of missing data, a complete case analysis was employed. The validity of this approach was confirmed by comparison of baseline characteristics between included and excluded patients, revealing no significant differences in key features.

Figure 1. The flowchart of study. MAP: mean arterial pressure.



Baseline Characteristics

Among the 34,981 included patients, 20,171 (57.66%) were men. The patients age was 66.67 (SD 16.01) years and the overall incidence of death within 28 days was 16.27% (5691/34,981). The patients' baseline characteristics are summarized in Table 1. Patients were divided into 4 groups based on MAP at admission. The distribution of intravenous immunoglobulin use did not differ significantly among the MAP subgroups (P=.09). Compared with the other three groups (Q2-Q4), patients in the Q1 group were older, had higher age, Charlson Comorbidity Index, SOFA scores, APACHE III scores, and lactate levels. They also had a higher frequency of renal replacement therapy, and use of norepinephrine, dopamine, dobutamine, cortisone, and antibiotics such as carbapenem, cephalosporin, and vancomycin. These characteristics, which are associated with poor prognosis, explain why patients in the Q1 group had the highest 28-day mortality rate. In contrast, as MAP increased from Q1 to Q4, age of the patient, severity of the disease, and the proportion of patients using relevant medications progressively decreased between groups. It may be a potential mechanism for the observed differences in 28-day mortality risk across MAP levels.

Table . Baseline characteristics of patients with sepsis in the MIMIC-IV database, 2008 - 2019.

Variables	Q1[a] (n=8745)	Q2[a] (n=8744)	Q3[a] (n=8745)	Q4[a] (n=8747)	*P* value
MAP[b] range (mm Hg)	(34.05 - 69.34)	(69.34 - 74.94)	(74.95 - 81.87)	(81.87 - 159.47)	_[c]
Age (years), mean (SD)	71.72 (15.36)	69.76 (15.09)	66.92 (15.33)	64.67 (16.39)	<.001
Males, n (%)	4687 (53.60)	5178 (59.22)	5180 (59.23)	5126 (58.60)	<.001
White population, n (%)	6350 (72.61)	6166 (70.52)	5860 (67.01)	5268 (60.23)	<.001
Charlson comorbidity index, mean (SD)	6.86 (2.83)	6.11 (2.87)	5.87 (2.92)	5.62 (3)	<.001
SOFA[e], mean (SD)	7.47 (4.02)	6.88 (3.79)	6.37 (3.56)	5.52 (3.22)	<.001
APACHE III[d], mean (SD)	81.43 (27.98)	72.11 (27.49)	67.21 (26.55)	64.24 (24.95)	<.001
Lactate (mmol/L), mean (SD)	3.55 (3.46)	3.14 (2.78)	3.00 (2.75)	2.84 (2.56)	<.001
Heart rate (bpm), mean (SD)	100.76 (25.11)	103.01 (24.08)	104.63 (24.15)	107.46 (24.23)	<.001
Respiratory rate (bpm), mean (SD)	26.67 (9.60)	26.57 (9.67)	26.49 (9.62)	27.08 (9.71)	<.001
Temperature (°C), mean (SD)	36.68 (1.28)	36.73 (1.30)	36.77 (1.25)	36.92 (1.25)	<.001
Dexamethasone, n (%)					<.001
No	7907 (90.42)	7917 (90.54)	7841 (89.66)	7497 (85.71)	
Yes	838 (9.58)	827 (9.46)	904 (10.34)	1250 (14.29)	
Methylprednisolone, n (%)					<.001
No	7297 (83.44)	7322 (83.74)	7315 (83.65)	6909 (78.99)	
Yes	1448 (16.56)	1422 (16.26)	1430 (16.35)	1838 (21.01)	
Cortisone, n (%)					.002
No	8526 (97.50)	8557 (97.86)	8577 (98.08)	8596 (98.27)	
Yes	219 (2.50)	187 (2.14)	168 (1.92)	151 (1.73)	
Norepinephrine, n (%)					<.001
No	4868 (55.67)	5520 (63.13)	6274 (71.74)	7240 (82.77)	
Yes	3877 (44.33)	3224 (36.87)	2471 (28.26)	1507 (17.23)	
Dopamine, n (%)					<.001
No	7783 (89)	8155 (93.26)	8223 (94.03)	8364 (95.62)	
Yes	962 (11)	589 (6.74)	522 (5.97)	383 (4.38)	
Dobutamine, n (%)					<.001
No	8241 (94.24)	8363 (95.64)	8377 (95.79)	8504 (97.22)	
Yes	504 (5.76)	381 (4.36)	368 (4.21)	243 (2.78)	
IVIG[f], n (%)					.09
No	8539 (97.64)	8538 (97.64)	8524 (97.47)	8495 (97.12)	
Yes	206 (2.36)	206 (2.36)	221 (2.53)	252 (2.88)	
MV[g], n (%)					<.001
No	5335 (61.01)	4517 (51.66)	4502 (51.48)	5222 (59.70)	

Variables	Q1[a] (n=8745)	Q2[a] (n=8744)	Q3[a] (n=8745)	Q4[a] (n=8747)	*P* value
Yes	3410 (38.99)	4227 (48.34)	4243 (48.52)	3525 (40.30)	
RRT[h], n (%)					<.001
No	7879 (90.10）	8232 (94.14）	8281 (94.69）	8225 (94.03）	
Yes	866 (9.90）	512 (5.86）	464 (5.31）	522 (5.97）	
Carbapenem, n (%)					<.001
No	6458 (73.85)	6883 (78.72)	7053 (80.65)	7138 (81.61)	
Yes	2287 (26.15)	1861 (21.28)	1692 (19.35)	1609 (18.39)	
Cephalosporin, n (%)					<.001
No	7892 (90.25)	8008 (91.58)	8007 (91.56)	8046 (91.99)	
Yes	853 (9.75)	736 (8.42)	738 (8.44)	701 (8.01)	
Penicillin, n (%)					<.001
No	3864 (44.19)	4381 (50.10)	4462 (51.02)	4200 (48.02)	
Yes	4881 (55.81)	4363 (49.90)	4283 (48.98)	4547 (51.98)	
Vancomycin, n (%)					<.001
No	1127 (12.89)	1662 (19.01)	1861 (21.28)	1909 (21.82)	
Yes	7618 (87.11)	7082 (80.99)	6884 (78.72)	6838 (78.18)	
28-day mortality, n (%)					<.001
No	6687 (76.47)	7452 (85.22)	7564 (86.50)	7587 (86.74)	
Yes	2058 (23.53)	1292 (14.78)	1181 (13.50)	1160 (13.26)	

[a]Patients were divided into 4 groups (Q1, Q2, Q3, and Q4) based on the mean MAP, using the quadruple classification approach.

[b]MAP: mean arterial pressure.

[c]Not applicable

[d]APACHE III: Acute Physiology and Chronic Health Evaluation III.

[e]SOFA: Sequential Organ Failure Assessment.

[f]IVIG: intravenous immunoglobulin.

[g]MV: mechanical ventilation.

[h]RRT: renal replacement therapy.

Relationship Between Mean Arterial Pressure and 28-Day Mortality

Univariate and Multivariate Analysis

We conducted univariate and multivariate analyses to investigate the relationship between MAP and 28-day mortality in patients with sepsis. Table 2 displays the results of the association between MAP and mortality in sepsis using various covariate adjustment strategies. In the nonadjusted model 1, each 1 mm Hg increase in MAP reduced the risk of 28-day mortality by 3% (odds ratio [OR] 0.97, 95% CI 0.971-0.974; *P*<.001). After adjusting for demographic factors such as gender, age, and White race, model 2 showed that the trend of OR was not altered (OR 0.97, 95% CI 0.97-0.98; *P*<.001). However, after adjusting for all covariables presented in Table 1, model 3 demonstrated that each 1 mm Hg increase in MAP resulted in a 1% decrease in the risk of 28 day-mortality (OR 0.99, 95% CI 0.99-1.00; *P*<.001). In model 3, the association between MAP and the probability of death in 28 days reduced to 1%, which was lower than the 3% decline seen in Models 1 and 2. This distinction was further explored and analyzed; we screened for covariates and discovered that SOFA score, age, and the Charlson comorbidity index had the greatest influence on the outcomes. The methodological rationale for this analysis was to include or exclude the effect of covariates on the regression coefficients of the main independent variables when constructing regression models. For example, in the unadjusted model, we included the main independent variable, MAP and found that the regression coefficient with the outcome was −0.0304. However, in the fully adjusted model, we added a number of covariates such as age, Charlson Comorbidity Index, SOFA score, and heart rate, the regression coefficient of MAP decreased to −0.0072 when these covariates were included (Multimedia Appendix 1). This implies that these factors “explain” or “mediate” the relationship between MAP and outcome to some extent. Specifically, age, Charlson index, and SOFA scores, which might reflect the patients’ baseline status and illness severity, contributed to the influence of MAP on outcome. This methodological aspect, namely the change in regression coefficients of the independent variables observed between the basic and full models, allows us to gain a deeper understanding of the complex link between independent factors and outcomes. It can help us uncover potential interactions or mediating effects that give a foundation

for understanding study findings. In addition, we performed a sensitivity analysis to verify the robustness of our findings by converting the MAP from a continuous variable into categorical variables (quartiles) and performing a trend test. The Q1 MAP was used as a reference; therefore, the same association was observed (P<.001) (Table 2).

Table . Relationship between mean arterial pressure and 28-day mortality.

Exposure	Model 1[a] (OR, 95% CI)	Model 2[b] (OR, 95% CI)	Model 3[c] (OR, 95% CI)	*P* value
MAP[d] (mm Hg)	0.97 (0.97-0.97)	0.97 (0.97-0.98)	0.99 (0.99-1.00)	<.001
MAP[e] quartiles (mm Hg)				
<65	1	1	1	
65 - 70	0.60 (0.55,-0.66)	0.61 (0.56 -0.67)	0.65 (0.59- 0.73)	<.001
70 - 80	0.41 (0.38-0.45)	0.42 (0.39- 0.46)	0.51 (0.46- 0.56)	<.001
80 - 85	0.38 (0.34-0.43)	0.40 (0.36- 0.45)	0.58 (0.51- 0.66)	<.001
≥85	0.38 (0.35- 0.42)	0.42 (0.38-0.47)	0.74 (0.65- 0.83)	<.001
P value for trend	<.001	<.001	<.001	_[f]

[a]Model 1 unadjusted.

[b]Model 2 adjusted for gender, age at admission, White race.

[c]Model 3 adjusted for gender, age at admission, White race.

[d]MAP as a continuous variable.

[e]MAP as a categorical variable.

[f]Not applicable.

Nonlinear Association Between MAP and 28-Day Mortality

We used smoothed curve fitting and GAM to evaluate the nonlinear relationship between MAP and 28-day mortality. After adjusting for all covariables listed in Table 1, our results demonstrated a U-shaped association between MAP and 28-day mortality with both low and high MAP associated with an increased risk of 28-day mortality (Figure 2). Furthermore, we employed a two-piecewise linear regression model and a recursive algorithm to determine the inflection points of MAP. The inflection points refer to points on the curve, where the curve transitions from falling to rising or rising to falling. Identifying these inflection points can help us better understand the complex relationship between MAP and 28-day mortality, as a single linear model may not fully capture this nonlinear relationship. We calculated the low and high inflection points at 70 and 82 mm Hg, respectively. When MAP was between 34.05 mm Hg and 70 mm Hg, the risk of 28-day mortality decreased by 7% for every 1 mm Hg increase (OR 0.93, 95% CI 0.92-0.94; P<.001). When the MAP was 70-82 mm Hg, there was no significant correlation (OR 1.01, 95% CI 0.99-1.02; P=.28). However, when MAP increased from 82 mm Hg to 159.47 mm Hg, the probability of 28-day mortality increased by 1% for every 1 mm Hg rise (OR 1.01, 95% CI 1.01-1.02, P=.002) (Table 3).

Figure 2. The nonlinear relationship between MAP and 28-day mortality risk. The solid line represents the smoothed curve fit while the dotted line represents 95% CI. MAP: mean arterial pressure.

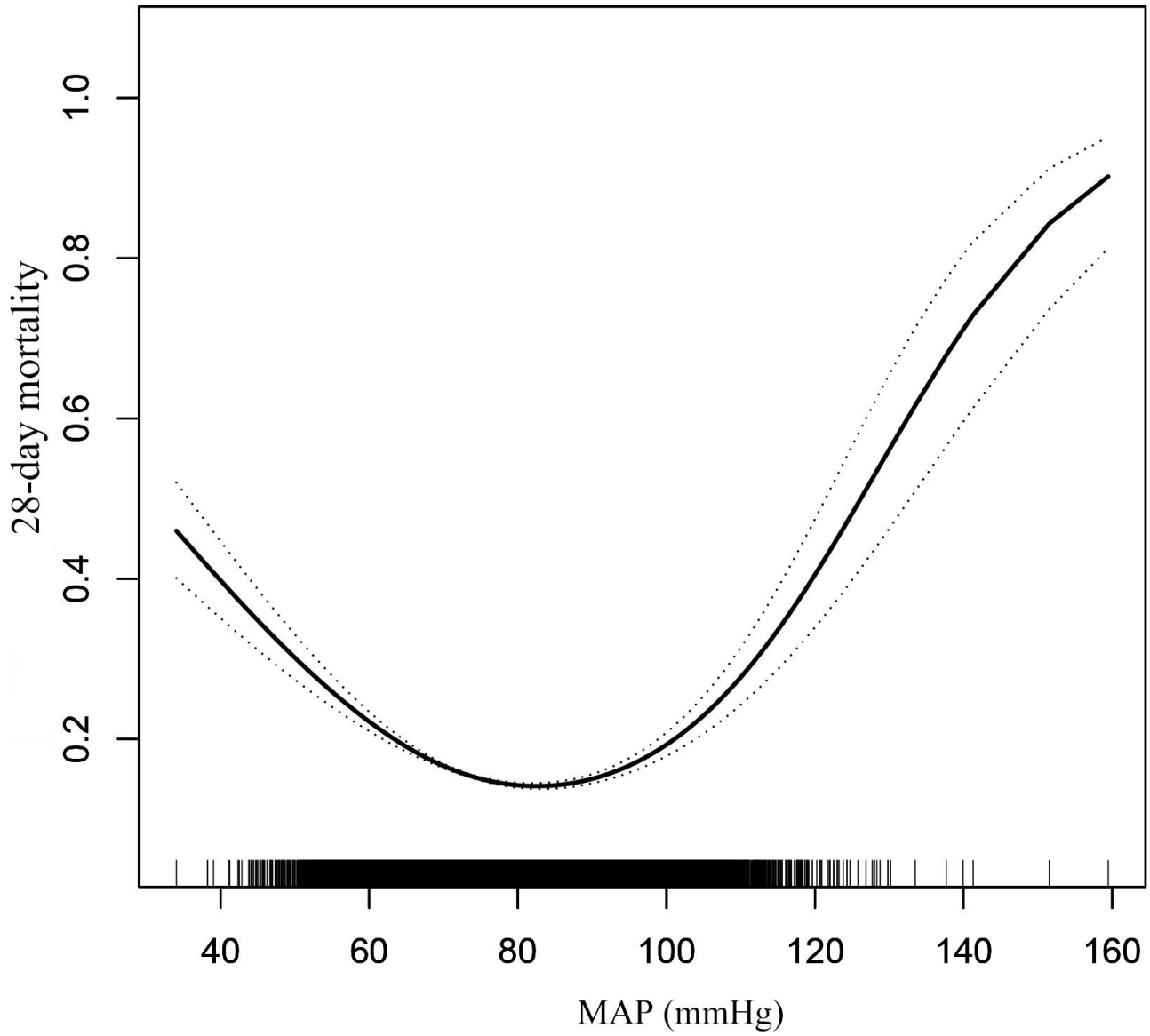


Table . Threshold effect analysis for the relationship between mean arterial pressure and 28-day mortality.

Outcomes	28-day mortality	
	OR (95%CI)	*P* value
Fitting model using logistic regression model	0.99 (0.99-1.00)	<.001
Fitting model using two-piecewise linear model		
Inflection point (mm Hg)		
<70	0.93 (0.92-0.94)	<.001
70 - 82	1.01 (0.99-1.02)	.28
>82	1.01 (1.01-1.02)	.002
Log-likelihood ratio test		<.001

Subgroup Analysis

The interaction analysis revealed significant effect modification by hypertension status (*P* for interaction=.03) and heart failure status (*P* for interaction=.03) on the association between MAP and 28-day mortality (Table 4). In hypertensive patients, each 1 mm Hg increase in MAP was associated with a slightly stronger protective effect against mortality (OR 0.993, 95% CI 0.992 - 0.994) compared to nonhypertensive patients (OR 0.984, 95% CI 0.982 - 0.986). Conversely, in patients with heart failure, the protective effect was marginally attenuated (OR 0.989, 95% CI 0.988 - 0.990) compared to those without heart

failure (OR 0.995, 95% CI 0.992 - 0.997). The use of vasopressors did not significantly modify the association between MAP and mortality (*P* for interaction=.14), with similar protective effects observed in both vasopressor users (OR 0.988, 95% CI 0.984 - 0.992) and nonusers (OR 0.991, 95% CI 0.990 - 0.992). These findings suggest that the optimal blood pressure management strategy may need to be tailored according to patients' comorbidity profile, particularly in those with hypertension or heart failure.

Table . Interaction analysis between mean blood pressure and treatment on 28-day mortality.

Model	28-day mortality, odds ratio (95% CI)	*P* value for interaction
Hypertension[a]		.03
No	0.984 (0.982-0.986)	
Yes	0.993 (0.992-0.994)	
Vasopressor[b]		.14
No	0.991 (0.990-0.992)	
Yes	0.988 (0.984-0.992)	
Heart failure[c]		.03
No	0.995 (0.992-0.997)	
Yes	0.989 (0.988-0.990)	

[a]Model Hypertension adjusted for: gender; age at admission; ethnicity; Charlson Comorbidity Index; dexamethasone; dopamine; dobutamine; IVIG, intravenous immunoglobulins; methylprednisolone; MV, mechanical ventilation; cortisone; SOFA, Sequential Organ Failure Assessment; glucocorticoids; cephalosporin; penicillin; heart rate; respiratory rate; vancomycin; body temperature and the interaction terms for following variables: dopamine, ethnicity, mechanical ventilation, SOFA.

[b]Model Vasopressor adjusted for: gender; age at admission; ethnicity; Charlson Comorbidity Index; dexamethasone; IVIG, intravenous immunoglobulins; methylprednisolone; MV, mechanical ventilation; cortisone; SOFA, Sequential Organ Failure Assessment; glucocorticoids; cephalosporin; penicillin; heart rate; respiratory rate; vancomycin; body temperature and the interaction terms for following variables: age at admission.

[c]Model Heart failure adjusted for: gender; age at admission; ethnicity; Charlson Comorbidity Index; dexamethasone; IVIG, intravenous immunoglobulins; methylprednisolone; MV, mechanical ventilation; cortisone; SOFA, Sequential Organ Failure Assessment; glucocorticoids; cephalosporin; penicillin; respiratory rate; vancomycin; body temperature and the interaction terms for following variables: age at admission, dexamethasone, IVIG, SOFA, glucocorticoids, cephalosporin, vancomycin, body temperature.

Sensitivity Analysis

Our sensitivity trials yielded results that confirmed the robustness of our findings (Multimedia Appendix 2). The regression coefficient for MAP was –0.0068 (95% CI –0.0123 to –0.0013) when the endpoint was 30-day mortality. When the median was used, the regression coefficient for MAP was –0.0065 (95% CI –0.0118 to –0.0012). Following the exclusion of individuals receiving medication for hypertension, the MAP regression coefficient was –0.0070 (95% CI –0.0128 to –0.0012). Other covariate selection methods produced results that were comparable. These results support our main finding, which confirms that the chance of patient mortality is inversely correlated with the MAP in all conditions.

Discussion

Main Findings

We investigated the relationship between MAP and the 28-day risk of mortality in patients with sepsis using data from a large, multicenter retrospective cohort study in the United States. We found a nonlinear association between MAP and 28-day mortality using a GAM and a two-piecewise linear regression. The findings revealed a U-shaped connection, indicating that both high and low MAP was associated with an increased risk of 28-day mortality in patients with sepsis. We discovered that each 1 mm Hg increase in MAP below 70 mm Hg resulted in a 7% decrease in the 28-day risk of mortality. When MAP exceeded 82 mm Hg, the 28-day mortality risk increased by 1% for each 1 mm Hg rise. This implies that maintaining MAP between 70-82 mm Hg may help improve the outcome in individuals with sepsis. While previous studies, including the work by Zhong et al [29] have explored the relationship between MAP and sepsis outcomes using MIMIC-III data, our study extends these findings in several important ways. First, we used the updated MIMIC-IV database with a larger sample size (34,981 vs. 14,031), providing more statistical power. Second, we employed more sophisticated statistical methods, including GAM models and two-piecewise linear regression, which allowed us to identify specific inflection points. Most importantly, our study identified a precise optimal MAP range (70-82 mm Hg) for clinical practice, whereas previous studies were unable to determine such specific thresholds. These differences explain why our findings provide more detailed and clinically applicable guidance for MAP management in patients with sepsis. There is no substantial relationship between MAP and 28-day mortality in this range, and organ damage from hypoperfusion or hyper-perfusion can be avoided. This finding provides clinicians with a reference range that can help guide blood pressure management in patients with sepsis. Further, more clinical studies are needed to validate these findings. However, this study provides a valuable reference for optimizing blood pressure management in patients with sepsis.

Most previous studies have explored the effect of low MAP on the outcome in patients with sepsis; our study further discovered a U-shaped relationship with specific inflection points. Our findings regarding the lower MAP threshold (70 mm Hg) are consistent with several previous studies. In a single-center retrospective cohort study of 1395 patients with severe sepsis or septic shock, []the 28-day mortality rate with an average MAP<65 mm Hg was 39.7%, which was significantly higher than other groups [30]. Similarly, in another retrospective cohort study based on the MIMIC III database by Cao et al. [31], which included 14,607 patients with sepsis, it was found that a lower MAP was significantly associated with higher 30-day mortality rates, with the lowest inflection point at 68.6 mm Hg.

However, our study differs from previous research in several important aspects. First, we used more sophisticated statistical methods including GAM models and two-piecewise linear regression, enabling us to identify both lower and upper thresholds. Although previous studies comparing the effects of higher and lower MAP on sepsis outcomes have yielded negative results, our findings uniquely identified 82 mm Hg as a critical upper threshold, above which there is an increased risk of death [17,18]. This increased risk could be due to secondary damage caused by excessive workload on organs and ischemia-reperfusion injury [31,32].

We speculate that the disparities between our findings and previous studies can be attributed to several factors: (1) the populations studied were not identical; (2) we provided more comprehensive covariate adjustment, whereas previous studies did not adjust for potential confounders; and (3) our analytical approach using the GAM model allowed us to detect nonlinear associations that might have been overlooked in previous studies. Further research is needed to determine whether these findings can be applied across populations.

Strengths and Limitations

This study has several key advantages. First, we used data from the updated MIMIC-IV cohort, which provides a larger sample size and, consequently greater statistical power. Second, we adjusted for a broader set of covariate indicators and focused on treatment strategies that were more closely related to patient outcomes. To better determine the true relationship between MAP and mortality, we used a GAM and two-piecewise linear models, which are advanced algorithms. Furthermore, we calculated the U-curve and identified two inflection points that provided a safe MAP range associated with lower mortality risk. Third, as this study used a large number of sensitivity analyses, the results are more robust. More informative, supporting evidence for MAP monitoring and clinical decision-making based on this metric is provided through this study.

Several limitations of this study should be noted. First, our findings are based solely on the MIMIC-IV database, which predominantly includes patients from a single geographic region and may not be representative of other populations. Given potential racial and ethnic differences in cardiovascular responses and outcomes, external validation of our findings in different 17 populations, particularly among Asian populations including Chinese patients, would be valuable. Such validation studies could help establish whether the optimal MAP range identified in our study is universally applicable or needs to be adjusted for different populations. Second, while we adjusted for measurable confounders, we could not account for unmeasurable confounding. Additionally, due to substantial missing data in lactate measurements, we were unable to conduct subgroup analyses based on lactate levels. This limitation prevents a deeper understanding of how the optimal MAP targets might differ among patients with varying degrees of tissue hypoperfusion, as indicated by lactate levels. Third, as this was an observational study, and therefore, inherently constrained, we could only observe relationships rather than evaluate causality.

Conclusion

There is a U-shaped association between MAP and the mortality risk in patients with sepsis. Both increases or decreases in MAP are linked to increased mortality risk. Our findings suggest that patients with sepsis have a lower risk of death when their MAP is maintained between 70-82 mm Hg.

Acknowledgments

The manuscript's English language quality was enhanced using AI-assisted proofreading tools, while maintaining the original content and meaning. This study was supported by the grants from National Natural Science Foundation of China [82160365]. This manuscript utilized AI tools exclusively for language enhancement and proofreading purposes. The intellectual content, research design, analyses, and conclusions were entirely conceived and executed by the authors.

Additionally, we confirm that one of our co-authors, LC, has completed the required human subjects training and obtained access to the MIMIC-IV database by signing the Data Use Agreement.

Data Availability

The data are available on the MIMIC-IV database website [33].

Authors' Contributions

Conceptualization: FS, QC
Data curation: YW
Formal analysis: LC, CX, DH, JY, ML, QC, QL, WL, XC

Methodology: QC
Supervision: FS
Writing – original draft: QC
Writing – review & editing: FS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics of patients with sepsis in the MIMIC-Ⅳ database, 2008-2019.
[DOCX File, 23 KB - i-jmr_v14i1e63291_app1.docx]

Multimedia Appendix 2
Relationship between mean arterial pressure and 28-day mortality.
[DOCX File, 17 KB - i-jmr_v14i1e63291_app2.docx]

References

1. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810. [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):775-787. [doi: 10.1001/jama.2016.0289] [Medline: 26903336]
3. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):762-774. [doi: 10.1001/jama.2016.0288] [Medline: 26903335]
4. Vincent JL. Highlighting the huge global burden of sepsis. Anaesth Crit Care Pain Med 2020 Apr;39(2):171-172. [doi: 10.1016/j.accpm.2020.03.004] [Medline: 32145433]
5. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. JAMA 2017 Oct 3;318(13):1241-1249. [doi: 10.1001/jama.2017.13836] [Medline: 28903154]
6. Prest J, Sathananthan M, Jeganathan N. Current trends in sepsis-related mortality in the United States. Crit Care Med 2021 Aug 1;49(8):1276-1284. [doi: 10.1097/CCM.0000000000005017] [Medline: 34261926]
7. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. The Lancet 2020 Jan;395(10219):200-211. [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31911147]
8. Torio CM, Moore BJ. National inpatient hospital costs: the most expensive conditions by payer 2013. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs: Agency for Healthcare Research and Quality; 2016. [Medline: 27359025]
9. Chiu C, Legrand M. Epidemiology of sepsis and septic shock. Curr Opin Anaesthesiol 2021 Apr 1;34(2):71-76. [doi: 10.1097/ACO.0000000000000958] [Medline: 33492864]
10. Vincent JL, Marshall JC, Namendys-Silva SA, et al. Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. Lancet Respir Med 2014 May;2(5):380-386. [doi: 10.1016/S2213-2600(14)70061-X] [Medline: 24740011]
11. Wu M, Islam MM, Poly TN, Lin MC. Application of AI in sepsis: citation network analysis and evidence synthesis. Interact J Med Res 2024 Apr 15;13:e54490. [doi: 10.2196/54490] [Medline: 38621231]
12. Liang M, Xu Y, Ren X, Huang D, Jin M, Qiu Z. The U-shaped association between serum osmolality and 28-day mortality in patients with sepsis: a retrospective cohort study. Infection 2024 Oct;52(5):1931-1939. [doi: 10.1007/s15010-024-02256-3] [Medline: 38647828]
13. LeDoux D, Astiz ME, Carpati CM, Rackow EC. Effects of perfusion pressure on tissue perfusion in septic shock. Crit Care Med 2000 Aug;28(8):2729-2732. [doi: 10.1097/00003246-200008000-00007] [Medline: 10966242]
14. Rhodes A, Evans LE, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med 2017 Mar;43(3):304-377. [doi: 10.1007/s00134-017-4683-6] [Medline: 28101605]
15. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021 Nov;47(11):1181-1247. [doi: 10.1007/s00134-021-06506-y] [Medline: 34599691]
16. Vincent JL, Nielsen ND, Shapiro NI, et al. Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database. Ann Intensive Care 2018 Nov 8;8(1):107. [doi: 10.1186/s13613-018-0448-9] [Medline: 30411243]

17. Asfar P, Meziani F, Hamel JF, et al. High versus low blood-pressure target in patients with septic shock. N Engl J Med 2014 Apr 24;370(17):1583-1593. [doi: 10.1056/NEJMoa1312173] [Medline: 24635770]
18. Lamontagne F, Richards-Belle A, Thomas K, et al. Effect of reduced exposure to vasopressors on 90-day mortality in older critically ill patients with vasodilatory hypotension: a randomized clinical trial. JAMA 2020 Mar 10;323(10):938-949. [doi: 10.1001/jama.2020.0930] [Medline: 32049269]
19. Lamontagne F, Meade MO, Hébert PC, et al. Higher versus lower blood pressure targets for vasopressor therapy in shock: a multicentre pilot randomized controlled trial. Intensive Care Med 2016 Apr;42(4):542-550. [doi: 10.1007/s00134-016-4237-3] [Medline: 26891677]
20. Thooft A, Favory R, Salgado DR, et al. Effects of changes in arterial pressure on organ perfusion during septic shock. Crit Care Lond Engl 2011;15(5):R222. [doi: 10.1186/cc10462] [Medline: 21936903]
21. Hylands M, Moller MH, Asfar P, et al. A systematic review of vasopressor blood pressure targets in critically ill adults with hypotension. Can J Anesth/J Can Anesth 2017 Jul;64(7):703-715. [doi: 10.1007/s12630-017-0877-1]
22. Bourgoin A, Leone M, Delmas A, Garnier F, Albanèse J, Martin C. Increasing mean arterial pressure in patients with septic shock: effects on oxygen variables and renal function. Crit Care Med 2005 Apr;33(4):780-786. [doi: 10.1097/01.ccm.0000157788.20591.23] [Medline: 15818105]
23. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015 Oct;12(10):e1001885. [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
24. Roggeveen LF, Guo T, Fleuren LM, et al. Right dose, right now: bedside, real-time, data-driven, and personalised antibiotic dosing in critically ill patients with sepsis or septic shock-a two-centre randomised clinical trial. Crit Care 2022 Sep 5;26(1):265. [doi: 10.1186/s13054-022-04098-7] [Medline: 36064438]
25. Maheshwari K, Nathanson BH, Munson SH, et al. The relationship between ICU hypotension and in-hospital mortality and morbidity in septic patients. Intensive Care Med 2018 Jun;44(6):857-867. [doi: 10.1007/s00134-018-5218-5] [Medline: 29872882]
26. Burstein B, Tabi M, Barsness GW, Bell MR, Kashani K, Jentzer JC. Association between mean arterial pressure during the first 24 hours and hospital mortality in patients with cardiogenic shock. Crit Care 2020 Aug 20;24(1):513. [doi: 10.1186/s13054-020-03217-6] [Medline: 32819421]
27. EmpowerStats: simple & powerful tool for epidemiology and bio-statistics that helps you analyze data and create publishable result tables, all without writing code. EmpowerStats. URL: http://www.empowerstats.com [accessed 2022-08-16]
28. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 3;10(1):1. [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
29. Zhong X, Li H, Chen Q, et al. Association between different MAP levels and 30-day mortality in sepsis patients: a propensity-score-matched, retrospective cohort study. BMC Anesthesiol 2023 Apr 6;23(1):116. [doi: 10.1186/s12871-023-02047-7] [Medline: 37024806]
30. Lee GT, Hwang SY, Jo IJ, et al. Associations between mean arterial pressure and 28-day mortality according to the presence of hypertension or previous blood pressure level in critically ill sepsis patients. J Thorac Dis 2019 May;11(5):1980-1988. [doi: 10.21037/jtd.2019.04.108] [Medline: 31285891]
31. Cao B, Chen Q, Tang T, et al. Non-linear relationship between baseline mean arterial pressure and 30-day mortality in patients with sepsis: a retrospective cohort study based on the MIMIC-III database. Ann Transl Med 2022 Aug;10(16):872. [doi: 10.21037/atm-22-3457] [Medline: 36111019]
32. Cai J, Chen X, Liu X, et al. AMPK: The key to ischemia-reperfusion injury. J Cell Physiol 2022 Nov;237(11):4079-4096. [doi: 10.1002/jcp.30875] [Medline: 36134582]
33. Medical information mart for intensive care: freely available medical data for research. MIMIC. URL: https://physionet.org/content/mimiciv/2.1/ [accessed 2022-08-16]

Abbreviations

APACHE III: Acute Physiology and Chronic Health Evaluation III
ICU: intensive care units
MAP: mean arterial pressure
MIMIC-IV: Medical Information Mart for Intensive Care
SOFA: sequential Organ Failure Assessment

Edited by TDA Cardoso; submitted 15.06.24; peer-reviewed by Y Zhou, Z Zhang; revised version received 02.01.25; accepted 28.01.25; published 05.03.25.

Please cite as:
Chen Q, Li W, Wang Y, Chen X, He D, Liu M, Yuan J, Xiao C, Li Q, Chen L, Shen F
Investigating the Association Between Mean Arterial Pressure on 28-Day Mortality Risk in Patients With Sepsis: Retrospective Cohort Study Based on the MIMIC-IV Database
Interact J Med Res 2025;14:e63291
URL: https://www.i-jmr.org/2025/1/e63291
doi:10.2196/63291

Prevalence and Correlates of Clinically Elevated Depressive Symptoms in a Nationwide Sample of Transgender, Nonbinary, and Gender Diverse Young Adults in the United States: Cross-Sectional Survey Study

Sari Reisner[1,2,3], ScD, MA; Yuxin Liu[4], MSc; Regina Tham[4], BS; Kaiden Kane[4], BS; S Wilson Cole[4,5], MA, MPH, PhD; Elizabeth R Boskey[6,7], LICSW, MSW, MPH, PhD; Sabra L Katz-Wise[8,9,10], PhD; Alex S Keuroghlian[3,11,12], MPH, MD; Rena Xu[4,7], MBA, MD

[1]Department of Epidemiology, University of Michigan School of Public Health, SPH-I 2649A, 1415 Washington Heights, Ann Arbor, MI, United States

[10]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[11]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[12]Psychiatry, Harvard Medical School, Boston, MA, United States

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[3]The Fenway Institute, Fenway Health, Boston, MA, United States

[4]Department of Urology, Boston Children's Hospital, Boston, MA, United States

[5]Query Research Consulting, Glen Burnie, MD, United States

[6]Division of Gynecology, Boston Children’s Hospital, Boston, MA, United States

[7]Department of Surgery, Harvard Medical School, Boston, MA, United States

[8]Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, Boston, MA, United States

[9]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Sari Reisner, ScD, MA
Department of Epidemiology, University of Michigan School of Public Health, SPH-I 2649A, 1415 Washington Heights, Ann Arbor, MI, United States

Abstract

Background: In the United States, transgender, nonbinary, and gender diverse (TGD) young adults experience a higher risk of depression compared to their cisgender peers. Understanding factors associated with increased risk of depression within the TGD young adult population is important to guide clinical care as well as inform the development of interventions to reduce mental health disparities.

Objective: This exploratory study investigated the prevalence and correlates of positive screening for depressive symptoms among TGD young adults to inform the design, development, and implementation of national interventions aimed at improving mental health in this at-risk population.

Methods: In August 2022, a cross-sectional, nationwide online survey was conducted among TGD young adults aged 18 - 25 (N=104) in the United States. Measures included sociodemographic variables, family characteristics, mental health care utilization, and the two-item Patient Health Questionnaire-2 (PHQ-2) screener for depression. Poisson regression models with robust variance estimation were fitted to estimate adjusted prevalence ratios (aPR) and 95% CI for correlates of PHQ-2 depression (score ≥3).

Results: The study sample had a mean age of 22 (SD 2) years; 48/104 (46%) individuals identified as Black, Indigenous, or other People of Color, and 69/104 (66%) were nonbinary. Overall, 44 (42%) individuals screened positive for depression using PHQ-2. In a multivariable model adjusted for age, race and ethnicity, US census region, and health insurance status, factors associated with increased depression prevalence using PHQ-2 included low versus high family support (aPR 1.54, 95% CI 1.05 - 2.27) and identifying with a nonChristian religion versus being unaffiliated (aPR 1.66, 95% CI 1.04 - 2.63). Factors associated with reduced depression prevalence included living in a rural versus suburban area (aPR 0.48, 95% CI 0.26 - 0.92) and receiving mental health therapy versus not (aPR 0.71, 95% CI 0.53 - 0.97).

Conclusions: The high prevalence of depressive symptoms among TGD young adults in this study sample highlights the need for comprehensive mental health evaluation and support in this population. Depression risk is increased among certain subgroups, such as those with low family support. These findings are valuable in informing the development of interventions that aim to improve mental health outcomes among TGD young people.

(Interact J Med Res 2025;14:e66630) doi:10.2196/66630

KEYWORDS

transgender; depression; preventive screening; young adult; LGBTQ+; nonbinary; gender minority; gender diverse; mental health; prevalence; cross-sectional; survey; questionnaire; nationwide; USA; United States; North America

Introduction

In the United States, major depression is a leading cause of morbidity and mortality, affecting an estimated 21 million adults aged ≥18 years (8.3% of all US adults) [1,2]. It is most prevalent among young adults aged 18 - 25 (18.6%) [3]. The peak incidence of depression occurs between ages 12 - 25 years, and 1 in 5 adolescents (20.1%) aged 12 - 17 in the US has had at least one major depressive episode, highlighting the importance of early detection and intervention efforts [3,4]. The US Preventive Services Task Force recommends routine preventive screening for depression among the general adult population aged ≥18 years in primary care settings, using tools such as the Patient Health Questionnaire (PHQ) to identify individuals who should be further evaluated for depression and facilitate appropriate referrals to mental health services [5-7].

Transgender, nonbinary, and gender diverse (TGD) young adults in the United States have a two- to four-fold higher risk of major depression [8] and are more likely to screen positive for clinically elevated depressive symptoms relative to their cisgender counterparts [9]. For example, in a national random sample of 65,231 college students drawn from 71 US college campuses, the prevalence of past 2-week depression using PHQ-9 was 57.8% for TGD young adults compared to 28.4% for cisgender young adults [9]. Other studies report similarly high prevalence rates of past 2-week elevated depressive symptoms, which vary depending on the screening instrument used and specific age range sampled: 56.7% (PHQ-9 depression in a clinical TGD sample aged 13 - 20 y) [10], 57.9% (two-item Patient Health Questionnaire, PHQ-2 depression in a representative school-based sample of TGD 9th and 11th graders in Minnesota) [11], and 75% (PHQ-2 depression in nonprobability online sample of 5753 TGD young people aged 13 - 24 y) [12]. Additional research is needed to understand factors associated with screening positive for depression among TGD young people, particularly using brief instruments appropriate for universal screening in clinical settings. Identifying TGD young people at increased risk for depression may facilitate more effective clinical care and inform the development of interventions to address mental health disparities.

Sociodemographic differences in depression prevalence based on gender identity and race or ethnicity have been inconsistent across studies among TGD young people [13], although nonbinary individuals reported consistently high rates of depression [9]. Other sociodemographic factors, such as geographic region and urbanity, are also important to consider, particularly in the context of legislation restricting elements of care for TGD populations. Several studies document the adverse negative health impacts of antitransgender legislation, even among TGD people who do not live in states with active or pending legislation [14]. While family support is protective for TGD mental health [15,16], other family-related factors, such as religiosity, necessitate further exploration [11], particularly since TGD young people have reported mixed quantitative findings on religiosity and mental health [17], and attitudes toward TGD people may differ by religious affiliation [18,19]. Additionally, both current utilization of mental health services and a history of seeking resources or information about mental health services may be relevant for future intervention implementation for TGD young adults. Consistent with a transformative paradigm in mental health that emphasizes early prevention and intervention of emerging mental health disparities in individuals aged 12 - 25 years [20], patterns of adolescent help-seeking may persist into young adulthood. Some TGD young people may have sought mental health information, while others may not have, highlighting the need to better understand whether help-seeking behaviors differ between those with and without depression. This knowledge can inform targeted intervention strategies to reach those who may need help.

This exploratory study sought to investigate the prevalence and correlates of depression using PHQ-2 in a nationwide US sample of TGD young adults to understand factors associated with increased risk of depression among TGD young people and to inform the development of interventions aimed at reducing mental health disparities in this population.

Methods

Setting

In August 2022, a cross-sectional survey was conducted with a nationwide US sample of young adults recruited and enrolled through Prolific, a survey platform that hosts a diverse panel of adults aged ≥18 years. The survey was conducted as part of TransHealthGUIDE, a National Institutes of Health (NIH)-funded project (U01MH136558) aimed at designing and implementing strategies to address mental health disparities among TGD young adults. Early in-depth key informant interviews with TGD young adults, caregivers, and health care providers identified the need to better understand the context of depression among TGD young adults. Specifically, the interviews highlighted the need for further information on the characteristics of TGD young adults with and without depression, the role of family, mental health service utilization, peer support, and recalled help-seeking behaviors in adolescence. This exploratory self-reported survey was developed as an extension of the key informant interviews and was designed to expand upon and quantify each of these factors for intervention development.

Participants

Eligibility criteria included: age 18 - 25 years, TGD identity (self-identifying as TGD or having a gender identity different from sex assigned at birth), and US residency. The sample was

stratified to enroll approximately 50 Black, Indigenous, and People of Color (BIPOC) people and 50 White young people. A total of 104 individuals were surveyed; additional details on survey methodology can be found elsewhere [21].

Assessments

PHQ-2 Depression Status

The psychometrically validated PHQ-2 assessed the frequency of depressed mood and anhedonia over the past two weeks [22]. The two items were (1) "little interest or pleasure in doing things" and (2) "feeling down, depressed, or hopeless." The response options were 0=not at all, 1=several days, 2=more than half the days, 3=nearly every day. The two items were highly correlated (r=0.81; P<.001) and summed. Scores ranged from 0 - 6, and a score ≥3 was considered a positive screen for depression, based on a previously validated threshold for screening for major depression (sensitivity: 83%, specificity: 92%) [22].

Sociodemographic Variables, Family Context, Mental Health Services, Peer Support, and Mental Health Help-Seeking in Adolescence

Sociodemographic variables included age in years (continuous and categorized: 18 - 20, 21 - 23, 23 - 25), race or ethnicity (ie, BIPOC monoracial, BIPOC multiracial, White), gender identity (ie, transgender men, transgender women, nonbinary), US Census region (ie, Northeast, Midwest, South, West), geographic context (ie, urban, suburban, rural), health insurance status (ie, private, public, both, none), and self-identified religion (eg, Agnostic, Atheist, Baptist, Buddhist, Catholic, Christian, Evangelical Protestant, Hindu, Jehovah's Witness, Jewish, Latter-day Saint or Mormon, Mainline Protestant, Muslim, Orthodox Christian, Pagan, Sikh, Spiritual but not religious, Unitarian/Universalist, Unaffiliated, Other, not applicable; coded as Christian, not Christian, unaffiliated, more than one religion, or not applicable).

Family-related variables included family support, assessed using the question: "How much trouble did your parents or caregivers have accepting your gender identity?" Responses were coded as high family support (none/a little), low family support (a moderate amount/a lot), or not applicable (not applicable/they don't know I'm trans); and family religion was self-reported by the young adult (using the same categories as for self-identified religion and coded as Christian, not Christian, unaffiliated, more than one religion, or not applicable).

Participants were queried on mental health services and peer support, including receiving care from a mental health therapist (yes, no) and whether they had ever participated in a peer support group (yes, no). Recalled frequency of mental health help-seeking in adolescence was assessed using 0-to-4-point Likert scales: (1) "When you were a teenager, how often were you looking for information about finding mental health support or a therapist?" (0=never to 4=very often; categorized as never/occasionally versus sometimes/often/very often), (2) "How useful would it have been to have reliable information on finding mental health support or a therapist?" (0=not at all to 4=extremely; categorized as not at all/slightly vs moderately/very/extremely), and (3) "How interested were you in getting therapy related to your gender? (0=not at all to 4=extremely; categorized as not at all/slightly, moderately/very/extremely, or not applicable).

Study Size

This exploratory survey was designed to be complementary to key informant interviews conducted with TGD young adults, caregivers, and health care providers as part of the NIH-funded TransHealthGUIDE project. After an initial round of key informant interviews, the research team identified the need for feedback from a larger group of diverse TGD young adults, than those were accessible for the interviews, and designed this survey to capture additional feedback.

Data Analysis

Descriptive statistics (ie, frequencies, percentages, means, standard deviations) summarized variables for the overall sample and by PHQ-2 depression status. Bivariate tests (2-tailed t tests for continuous variables; χ^2 tests for binary or categorical variables) were used to compare participants with and without depression detected using PHQ-2. Poisson regression models with robust variance were fit to estimate prevalence ratios (PR) and 95% CI for correlates of depression [23]. Bivariate models were fit for each variable, followed by a single multivariable model with all variables to estimate adjusted prevalence ratios (aPR). Variables for modeling were selected *a priori* through a review of existing TGD research, conceptual salience, and findings from key informant interviews. Analyses were conducted using R software (version 4.4.3; R Foundation for Statistical Computing) [24] with statistical significance at P<.05 [24].

Ethical Considerations

After review, the study protocol was deemed exempt by the Boston Children's Hospital Institutional Review Board (IRB-P00043127). This study was performed in accordance with the ethical standards outlined in the 1964 Declaration of Helsinki and its subsequent amendments. Participants completed an informed consent process. This was an anonymized survey, and participants indicated consent by opting into the survey. Participants were compensated $3 for completing the survey in accordance with the standards of the Prolific survey platform.

Results

Study Sample

Sample characteristics are displayed in Table 1 for the total sample and stratified by PHQ-2 depression scores. Overall, 44 of 104 (42%) of the sample had clinically elevated depressive symptoms.

Table . Characteristics of a US nationwide sample of Transgender, Nonbinary, and Gender Diverse young adults aged 18 - 25 years, stratified by Patient Health Questionnaire-2 (PHQ-2)[a] depression[b] .

Variables	Participants (N=104)	Depression[b] (n=44), n (%)	No depression[b] (n=60), n (%)	Bivariate test: *t* test (*df*) or χ^2 test (*df*)[c]	*P* value
Age (years), mean (SD)	21.8 (1.9)	22.1 (1.9)	21.6 (1.9)	–1.21 (90.2)[d]	.23
Age group(years), n (%)				2.09 (2)[e]	.35
18 - 20	28 (26.9)	10 (22.7)	18 (30)		
21 - 22	37 (35.6)	14 (31.8)	23 (38.3)		
23 - 25	39 (37.5)	20 (45.5)	19 (31.7)		
Race/ethnicity, n (%)				1.76 (2)[e]	.41
White	55 (52.9)	20 (45.5)	35 (58.3)		
BIPOC[f] monoracial	23 (22.1)	10 (22.7)	13 (21.7)		
BIPOC multiracial	25 (24)	13 (29.5)	12 (20)		
Missing	1 (1)	1 (2.3)	0 (0)		
Gender identity, n (%)				1.60 (2)[e]	.46
Nonbinary	69 (66.3)	29 (65.9)	40 (66.7)		
Transgender man	24 (23.1)	12 (27.3)	12 (20)		
Transgender woman	11 (10.6)	3 (6.8)	8 (13.3)		
US Census region, n (%)				7.69 (3)[e]	.05
Northeast	27 (26.0)	7 (15.9)	20 (33.3)		
Midwest	20 (19.2)	6 (13.6)	14 (23.3)		
South	34 (32.7)	19 (43.2)	15 (25)		
West	23 (22.1)	12 (27.3)	11 (18.3)		
Geography, n (%)				4.99 (3)[e]	.14
Urban	27 (26.0)	13 (29.5)	14 (23.3)		
Suburban	64 (61.5)	29 (65.9)	35 (58.3)		
Rural	10 (9.6)	1 (2.3)	9 (15.0)		
Unknown	3 (2.9)	1 (2.3)	2 (3.3)		
Health insurance, n (%)				6.86 (3)[e]	.08
Private	48 (46.1)	17 (38.6)	31 (51.7)		
Public	40 (38.5)	20 (45.5)	20 (33.3)		
Private and public	4 (3.8)	0 (0)	4 (6.7)		
No insurance	11 (10.6)	7 (15.9)	4 (6.7)		
Missing	1 (1)	0 (0)	1 (1.7)		
Own religion, n (%)[g]				1.75 (4)[e]	.81
Christian	7 (6.7)	2 (4.5)	5 (8.3)		
Not Christian	13 (12.5)	7 (15.9)	6 (10)		
Unaffiliated	68 (65.4)	29 (65.9)	39 (65)		
More than one	4 (3.8)	1 (2.3)	3 (5)		
Not applicable	12 (11.5)	5 (11.4)	7 (11.7)		

Variables	Participants (N=104)	Depression[b] (n=44), n (%)	No depression[b] (n=60), n (%)	Bivariate test: *t* test (*df*) or χ^2 test (*df*)[c]	*P* value
Family support, n (%)				4.18 (2)[e]	.12
High family support	24 (23.1)	6 (13.6)	18 (30)		
Low family support	31 (9.8)	16 (36.4)	15 (25)		
Not applicable/ They don't know I'm trans	49 (47.1)	22 (50)	27 (45)		
Family religion, n (%)[h]				7.35 (4)[e]	.11
Christian	55 (52.9)	30 (68.2)	25 (41.7)		
Not Christian	3 (2.9)	1 (2.3)	2 (3.3)		
Unaffiliated	19 (18.3)	6 (13.6)	13 (21.7)		
More than one religion	16 (15.4)	4 (9.1)	12 (20)		
Not applicable	11 (10.6)	3 (6.8)	8 (13.3)		
Mental health therapist, n (%)				1.25 (1)[e]	.26
No	56 (53.8)	27 (61.4)	29 (48.3)		
Yes	48 (46.2)	17 (38.6)	31 (51.7)		
Peer support group, n (%)				0.55 (1)[e]	.46
No	76 (73.1)	30 (68.2)	46 (76.7)		
Yes	28 (26.9)	14 (31.8)	14 (23.3)		
How often sought mental health/ therapy information in adults aged <18 years, n (%)				1.03 (1)[e]	.31
Never or occasionally	45 (43.3)	16 (36.4)	29 (48.3)		
Sometimes, often, or very often	59 (56.7)	28 (53.6)	31 (51.7)		
How useful would it have been to have reliable information on mental health/ therapy in adults aged <18 years, n (%)				0.49 (1)[e]	.49
Not at all or slightly	16 (15.4)	5 (11.4)	11 (18.3)		
Moderately, very, or extremely	88 (84.6)	39 (88.6)	49 (81.7)		
How interested you were in getting therapy related to gender (age <18 years), n (%)				1.78 (2)[e]	.40
Not at all or slightly	43 (41.3)	13 (29.5)	30 (50)		
Moderately, very, or extremely	53 (51)	28 (63.6)	25 (41.7)		

Variables	Participants (N=104)	Depression[b] (n=44), n (%)	No depression[b] (n=60), n (%)	Bivariate test: *t* test (*df*) or χ^2 test (*df*)[c]	*P* value
Not applicable	8 (7.7)	3 (6.8)	5 (8.3)		

[a]PHQ-2: Two-item Patient Health Questionnaire (score ≥3 is the validated cut-off and indicates clinically elevated depressive symptoms).

[b]Depression was assessed using the PHQ-2.

[c]Fisher's exact tests were used to obtain *P* values for cell sizes ≤5.

[d]*t* test (*df*).

[e]χ^2test (*df*).

[f]BIPOC: Black, Indigenous, and Other People of Color.

[g]Own Religion: Christian (57.1% Catholic, 14.3% Baptist, 14.3% Orthodox, 14.3% Christian); Not Christian (38.5% Pagan, 30.8% Jewish, 23.1% Buddhist, 7.7% Misotheist); Unaffiliated (41.9% Agnostic, 29.1% Atheist, 19.8% Spiritual, 9.3% Unaffiliated). Percentages may not sum to 100% due to rounding.

[h]Family Religion (self-reported by TGD young adults): Christian (49.2% Catholic, 20.6% Orthodox, 11.1% Mainline Protestant, 4.8% Evangelist, 3.2% Latter-day Saints, 3.2% Christian, 3.2% Baptist, 1.6% Muslim, 1.6% Jehovah's Witness, 1.6% Unitarian); Not Christian (66.7% Buddhist, 33.3% Jewish), Unaffiliated (39.3% Atheist, 28.6% Agnostic, 17.9% Spiritual, 14.3% Unaffiliated). Percentages may not sum to 100% due to rounding.

Participants' mean age was 21.8 (SD 1.9) years; a total of 25/104 (24%) were BIPOC multiracial individuals, while 23/104 (22%) were BIPOC monoracial individuals, and 69/104 (66%) participants identified as nonbinary. The majority (n=64, 61%) of participants resided in a suburban area, and 11/104 (11%) did not have health insurance. Regarding their religious affiliation, 68 (65%) participants described themselves as unaffiliated, while 55 (53%) described their family's religion as Christian. Half (49/104, 47%) of the participants were not out to their family, 31 (30%) reported low family support, while 24 (23%) participants reported high family support.

Among the participants, 56 (46%) had received mental health therapy and 28 of 104 (27%) had participated in a peer support group. When asked to recall their mental health needs during adolescence: 59 (57%) participants reported seeking information about mental health support or therapy, 88 (85%) indicated that reliable information on mental health support or therapy would have been useful, and 53 (51%) participants expressed having had an interest in therapy related to their gender.

Correlates of PHQ-2 Depression Score

In bivariate models (Table 2), increased PHQ-2 depression prevalence was associated with (1) being BIPOC multiracial versus White (PR1.39; 95% CI 1.05 - 1.84), (2) having low family support versus high (PR 1.53; 95% CI 1.09 - 2.15), (3) seeking information about mental health therapy in adolescence (PR 1.31; 95% CI 1.02 - 1.67), and (4) being interested in mental health therapy related to gender during adolescence (PR 1.41; 95% CI 1.09 - 2.05). Conversely, residing in a rural versus suburban area was associated with decreased PHQ-2 depression prevalence (PR 0.54; 95% CI 0.32 - 0.92). No other variables reached statistical significance in bivariate models.

Table . Bivariate and multivariable Poisson regression models with robust variance estimation: Correlates of Patient Health Questionnaire-2 (PHQ-2)[a] depression scores in a nationwide sample of transgender, nonbinary, and gender diverse young adults aged 18 - 25 across the United States.

Variables	Bivariate models (N=103)[b], PR[c] (95% CI)	*P* value	Multivariable model (N=103)[b], aPR[d] (95% CI)	*P* value
Age groups (years)				
18 - 20	Ref		Ref	
21 - 22	0.99 (0.72-1.35)	.94	0.83 (0.55-1.24)	.35
23 - 25	1.20 (0.89-1.61)	.24	1.03 (0.69-1.56)	.87
Race/Ethnicity				
White	Ref		Ref	
BIPOC[e] monoracial	1.10 (0.81-1.50)	.54	0.96 (0.63-1.45)	.84
BIPOC multiracial	1.39 (1.05-1.84)	.03	1.04 (0.75-1.44)	.81
Gender identity				
Nonbinary	Ref		Ref	
Transgender man	1.06 (0.80-1.40)	.70	0.91 (0.64-1.29)	.59
Transgender woman	0.71 (0.45-1.12)	.14	0.69 (0.41-1.16)	.16
US Census region				
Northeast	Ref		Ref	
Midwest	1.19 (0.82-1.73)	.36	1.16 (0.75-1.80)	.51
South	1.31 (0.94-1.80)	.11	1.02 (0.69-1.50)	.94
West	1.31 (0.92-1.87)	.13	1.14 (0.75-1.71)	.54
Geography				
Suburban	Ref		Ref	
Urban	0.96 (0.73-1.27)	.79	0.98 (0.72-1.36)	.93
Rural	0.54 (0.32-0.92)	.02	0.48 (0.26-0.92)	.03
Unknown	1.21 (0.64-2.28)	.57	1.16 (0.58-2.35)	.67
Health insurance				
Private	Ref		Ref	
Public	1.28 (0.99-1.65)	.07	1.08 (0.80-1.46)	.63
Private and public	0.86 (0.42-1.76)	.67	0.91 (0.40-2.05)	.82
No Insurance	1.36 (0.93-1.99)	.11	1.01 (0.63-1.62)	.97
Own religion				
Unaffiliated	Ref		Ref	
Not Christian	1.26 (0.90-1.76)	.19	1.66 (1.04-2.63)	.03
Christian	0.78 (0.45-1.34)	.36	0.76 (0.41-1.41)	.39
More than one religion	0.97 (0.51-1.84)	.93	0.78 (0.38-1.60)	.39
Not applicable	1.07 (0.74-1.55)	.73	1.25 (0.80-1.93)	.32
Family support				
High family support	Ref		Ref	
Low family support	1.53 (1.09-2.15)	.01	1.54 (1.05-2.27)	.03
Not applicable/ they don't know I'm trans	1.17 (0.85-1.62)	.34	1.06 (0.69-1.60)	.80
Family religion				
Unaffiliated	Ref		Ref	

Variables	Bivariate models (N=103)[b], PR[c] (95% CI)	*P* value	Multivariable model (N=103)[b], aPR[d] (95% CI)	*P* value
Not Christian	0.75 (0.32-1.74)	.50	1.61 (0.22-1.68)	.34
Christian	1.09 (0.80-1.49)	.59	1.05 (0.70-1.58)	.80
More than one religion	0.88 (0.58-1.35)	.57	0.86 (0.49-1.51)	.60
Not applicable	0.61 (0.36-1.04)	.07	0.65 (0.33-1.25)	.20
Mental health therapist				
No	Ref		Ref	
Yes	0.92 (0.73-1.17)	.51	0.71 (0.53-0.97)	.03
Peer support group				
No	Ref		Ref	
Yes	1.20 (0.93-1.55)	.16	0.97 (0.68-1.37)	.86
How often sought information on mental health/ therapy by individuals aged<18 years				
Never or occasionally	Ref		Ref	
Sometimes, often, or very often	1.31 (1.02-1.67)	.03	1.15 (0.79-1.66)	.46
How useful would It would have been to have reliable information on mental health/ therapy in individuals aged <18 years				
Not at all or slightly	Ref		Ref	
Moderately, very, or extremely	1.33 (0.92-1.91)	.13	1.18 (0.73-1.92)	.50
How interested you were in getting therapy related to gender in adults aged<18 years				
Not at all or slightly	Ref		Ref	
Moderately, very, or extremely	1.41 (1.09-1.82)	.01	1.47 (0.98-2.21)	.06
Not applicable	1.29 (0.81-2.05)	.29	1.33 (0.80-2.23)	.28

[a]PHQ-2 score ≥3 indicating clinically elevated depressive symptoms (validated cut-off).

[b]Models include N=103 participants; 1 participant was excluded due to missing data on health insurance.

[c]PR: prevalence ratio.

[d]aPR: adjusted prevalence ratio.

[e]BIPOC: Black, Indigenous, and Other People of Color.

In a multivariable model (Table 2), after adjusting for all other variables, identifying one's own religion as not Christian versus unaffiliated (aPR 1.66; 95% CI 1.04 - 2.63) and low family support (aPR 1.54; 95% CI 1.05 - 2.27) were associated with increased prevalence of positive depression screening. Residing in a rural versus suburban area (aPR 0.48; 95% CI 0.26 - 0.92) and ever receiving mental health therapy (aPR 0.71; 95% CI 0.53 - 0.97) were associated with decreased PHQ-2 depression prevalence in this model. Age, race or ethnicity, gender identity, US Census region, health insurance, family religion, peer support, history of interest in mental health therapy for gender during adolescence, frequency of information-seeking about mental health resources in adolescence, and desire for reliable mental health information during adolescence, were not significantly correlated with depression based on PHQ-2 scores.

Discussion

Principal Results

In this nationwide online study, depression identified using PHQ-2 scores was highly prevalent, with more than 4 in 10 TDG young adults showing clinically elevated symptoms. Low levels of family support and nonChristian religious affiliation were associated with an increased likelihood of positive PHQ-2 depression screening. This exploratory study highlights the high rates of depressive symptoms among TGD young adults as well as the need for tailored interventions to improve mental health among those at the highest risk.

Comparison With Prior Work

This study found that TGD young people with low family support had a higher prevalence of PHQ-2 depression compared to those with highly supportive families. This finding corroborates prior research on the protective role of family support on TGD young peoples' mental health and has informed intervention development for the broader TransHealthGUIDE

study [15,16]. For example, a key intervention component that has been developed for the overall study includes app-based education for TGD young adults and their caregivers on ways to improve family communication, connection, and support.

No differences in PHQ-2 depression prevalence were found for family religion, despite religious affiliation influencing attitudes toward TGD issues in the United States [18,19]. The TGD young people identifying their own religion as nonChristian (eg, Jewish, Muslim, Buddhist) had a higher prevalence of PHQ-2 depression than those identifying as unaffiliated; however, no significant differences were found between Christian and unaffiliated TGD young adults. The association between religious affiliation and depression has been inconsistent in prior research on young people [25], including a recent systematic review of religion, spirituality, and mental health among TGD youth [17]. Future mental health research may benefit from measuring not only religious affiliation but also the degree of religiosity, which was not assessed in this study.

This study found that a history of mental health help-seeking, specifically in young adults receiving mental health therapy was significantly associated with reduced PHQ-2 depression prevalence. This highlights the need for interventions that connect TGD young adults to necessary mental health services. The American Academy of Pediatrics recommends that adolescents aged 12 years and older be screened annually for depression using a formal self-report screening tool to identify and treat those with depression early in the course of illness, potentially reducing long-term depression-related morbidities [26]. This finding underscores the importance of improving mental health care access for TGD young adults to address depression, including telehealth and other online or digitally delivered services and interventions [27].

Regarding sociodemographic differences, we found no evidence of gender identity differences in PHQ-2 depression prevalence, inconsistent with prior research demonstrating the highest depression burden among nonbinary young adults [9]. However, previous studies have reported mixed findings regarding gender identity differences in depression [13], suggesting the need for ongoing research. While BIPOC multiracial TGD young people exhibited elevated PHQ-2 depression in bivariate analyses, consistent with national US population data, this finding did not remain significant in the adjusted model [3]. Geographic context emerged as a significant factor in this study, in alignment with psychiatric epidemiologic research showing lower depression rates in rural settings in high-income countries [28]. The TGD young adults residing in rural areas had lower PHQ-2 depression prevalence compared to those living in suburban areas. Prior studies have found greater mental distress symptoms among rural versus nonrural TGD adults but did not distinguish between the urban and suburban subgroups [29]. More recent research identified the lowest levels of depressive symptoms in rural versus urban-dwelling TGD adults [30]. Additional research is needed to better understand PHQ-2 depression prevalence and set of symptoms among TGD young adults in rural, urban, and suburban settings.

Limitations

These findings should be interpreted alongside this study's limitations. First, analyses were exploratory, and the small sample size limits statistical power to detect significant associations; the findings will require replication. Further, given the cross-sectional design, the study findings are associational only, and longitudinal research is recommended. Second, related to study measures, the global measure of family support may have overlooked important details about family dynamics, such as communication, acceptance, and explicit expressions of care and support [31]. Relatedly, mental health care utilization variables were self-reported rather than objectively measured. Future research that includes self-reported data alongside electronic health record data would strengthen the evidence base. Third, this study did not assess nonpsychiatric aspects of participants' medical histories, and unmeasured confounding due to gender-affirming medical treatments cannot be ruled out [10,12]. Lastly, there are few validated measures for assessing mental health help-seeking behaviors [32]. Further research to understand online help-seeking behaviors for TGD young people using validated measures is warranted to inform future interventions [33].

Conclusions

The high prevalence of PHQ-2 depression in this online sample of TGD young adults underscores the importance of universal screening for depression in this population. It also highlights the need for appropriate systems to ensure prompt diagnosis, treatment, and follow-up for TGD young adults with PHQ-2 depression. Further, tailored interventions may be necessary for subgroups at increased risk for depression, such as TGD young adults with low family support, to decrease mental health disparities. The continued expansion of online interventions may facilitate the engagement of TGD young adults who would benefit from mental health prevention and treatment efforts [27]. This study's findings have important implications for clinical care and the design, development, and implementation of interventions aimed at improving mental health outcomes and reducing health disparities among TGD young adults.

Acknowledgments

The authors wish to thank Jessica Kremen, MD, Meg Quint, BS, and the Community Advisory Board for their feedback on the survey questionnaire.

Research reported in this publication was supported by the National Institutes of Health (NIH) from the National Institute of Mental Health under award number U01MH136558 (MPIs: RX & SR) and the National Institute on Minority Health and Health Disparities of the NIH (R01MD015031; PI: SLK-W). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data Availability

Requests for the dataset and code for this study can be submitted to the corresponding author for consideration.

Authors' Contributions

Conceptualization: ERB, SR, RX
Data curation: ERB, SR, YL
Formal analysis: SR, YL
Funding acquisition: RX
Investigation: ERB, KK, RT, RX, SR
Project administration: KK, RT, RX, SR
Resources: RX, SR
Supervision: RX, SR
Validation: SR, YL,
Visualization: SR, YL
Writing – original draft: RX, SR, YL
Writing – review & editing: ASK, ERB, KK, MRG, RT, RX, SLK-W, SR, SWC, YL

Conflicts of Interest

SLR and ASK receive royalties from McGraw Hill for co-editing the textbook, “Transgender and Gender Diverse Health Care: The Fenway Guide.” ASK receives royalties from American Psychiatric Association Publishing for co-editing the textbook, “Gender-Affirming Psychiatric Care.” SWC consults for or have consulted in the recent past for several universities, nonprofit organizations, government organizations, and Viiv Healthcare, none of which have any involvement in this manuscript. SLK-W is a consultant for Paramount Global, who was not involved in the work reported in this manuscript.

References

1. Lee B, Wang Y, Carlson SA, et al. National, state-level, and county-level prevalence estimates of adults aged ≥18 years self-reporting a lifetime diagnosis of depression - United States, 2020. MMWR Morb Mortal Wkly Rep 2023 Jun 16;72(24):644-650. [doi: 10.15585/mmwr.mm7224a1] [Medline: 37318995]
2. Zhang Z, Jackson SL, Gillespie C, Merritt R, Yang Q. Depressive symptoms and mortality among US adults. JAMA Netw Open 2023 Oct 2;6(10):e2337011. [doi: 10.1001/jamanetworkopen.2023.37011] [Medline: 37812418]
3. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the united states: results from the 2021 national survey on drug use and health (HHS publication nopep22-07-01-005, NSDUH series H-57). : Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2022 URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt39443/2021NSDUHFFRRev010323.pdf [accessed 2025-03-21]
4. Solmi M, Radua J, Olivola M, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022 Jan;27(1):281-295. [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
5. Siu AL, US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, et al. Screening for depression in adults: US Preventive Services Task Force Recommendation Statement. JAMA 2016 Jan 26;315(4):380-387. [doi: 10.1001/jama.2015.18392] [Medline: 26813211]
6. US Preventive Services Task Force, Barry MJ, Nicholson WK, et al. Screening for depression and suicide risk in adults: US Preventive Services Task Force Recommendation Statement. JAMA 2023 Jun 20;329(23):2057-2067. [doi: 10.1001/jama.2023.9297] [Medline: 37338872]
7. Reynolds CF 3rd, Frank E. US Preventive Services Task Force Recommendation Statement on screening for depression in adults: not good enough. JAMA Psychiatry 2016 Mar;73(3):189-190. [doi: 10.1001/jamapsychiatry.2015.3281] [Medline: 26815331]
8. Reisner SL, Vetters R, Leclerc M, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. J Adolesc Health 2015 Mar;56(3):274-279. [doi: 10.1016/j.jadohealth.2014.10.264] [Medline: 25577670]
9. Lipson SK, Raifman J, Abelson S, Reisner SL. Gender minority mental health in the U.S.: results of a national survey on college campuses. Am J Prev Med 2019 Sep;57(3):293-301. [doi: 10.1016/j.amepre.2019.04.025] [Medline: 31427032]
10. Tordoff DM, Wanta JW, Collin A, Stepney C, Inwards-Breland DJ, Ahrens K. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. JAMA Netw Open 2022 Feb 1;5(2):e220978. [doi: 10.1001/jamanetworkopen.2022.0978] [Medline: 35212746]
11. Eisenberg ME, Gower AL, McMorris BJ, Rider GN, Shea G, Coleman E. Risk and protective factors in the lives of transgender/gender nonconforming adolescents. J Adolesc Health 2017 Oct;61(4):521-526. [doi: 10.1016/j.jadohealth.2017.04.014] [Medline: 28736148]

12. Green AE, DeChants JP, Price MN, Davis CK. Association of gender-affirming hormone therapy with depression, thoughts of suicide, and attempted suicide among transgender and nonbinary youth. J Adolesc Health 2022 Apr;70(4):643-649. [doi: 10.1016/j.jadohealth.2021.10.036] [Medline: 34920935]
13. Wittlin NM, Kuper LE, Olson KR. Mental health of transgender and gender diverse youth. Annu Rev Clin Psychol 2023 May 9;19:207-232. [doi: 10.1146/annurev-clinpsy-072220-020326] [Medline: 36608332]
14. Dhanani LY, Totton RR. Have you heard the news? The effects of exposure to news about recent transgender legislation on transgender youth and young adults. Sex Res Social Policy 2023 Apr 5;2023:1-15. [doi: 10.1007/s13178-023-00810-6] [Medline: 37363351]
15. Simons L, Schrager SM, Clark LF, Belzer M, Olson J. Parental support and mental health among transgender adolescents. J Adolesc Health 2013 Dec;53(6):791-793. [doi: 10.1016/j.jadohealth.2013.07.019] [Medline: 24012067]
16. Pariseau EM, Chevalier L, Long KA, Clapham R, Edwards-Leeper L, Tishelman AC. The relationship between family acceptance-rejection and transgender youth psychosocial functioning. Clin Pract Pediatr Psychol 2019 Sep;7(3):267-277. [doi: 10.1037/cpp0000291]
17. Lekwauwa R, Funaro MC, Doolittle B. Systematic review: The relationship between religion, spirituality and mental health in adolescents who identify as transgender. J Gay Lesbian Ment Health 2023 Oct 2;27(4):421-438. [doi: 10.1080/19359705.2022.2107592]
18. Attitudes about transgender issues vary widely among Christians, religious 'nones' in US. Pew Research Center. 2022. URL: https://www.pewresearch.org/short-reads/2022/07/07/attitudes-about-transgender-issues-vary-widely-among-christians-religious-nones-in-u-s/
19. Campbell M, Hinton JDX, Anderson JR. A systematic review of the relationship between religion and attitudes toward transgender and gender-variant people. Int J Transgend 2019;20(1):21-38. [doi: 10.1080/15532739.2018.1545149] [Medline: 32999592]
20. Uhlhaas PJ, Davey CG, Mehta UM, et al. Towards a youth mental health paradigm: a perspective and roadmap. Mol Psychiatry 2023 Aug;28(8):3171-3181. [doi: 10.1038/s41380-023-02202-z] [Medline: 37580524]
21. Boskey ER, Quint M, Xu R, et al. Gender affirmation-related information-seeking behaviors in a diverse sample of transgender and gender-diverse young adults: survey study. JMIR Form Res 2023 Aug 15;7:e45952. [doi: 10.2196/45952] [Medline: 37581925]
22. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
23. Petersen MR, Deddens JA. A comparison of two methods for estimating prevalence ratios. BMC Med Res Methodol 2008 Feb 28;8:9. [doi: 10.1186/1471-2288-8-9] [Medline: 18307814]
24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2024. URL: https://www.R-project.org
25. Aggarwal S, Wright J, Morgan A, Patton G, Reavley N. Religiosity and spirituality in the prevention and management of depression and anxiety in young people: a systematic review and meta-analysis. BMC Psychiatry 2023 Oct 10;23(1):729. [doi: 10.1186/s12888-023-05091-2] [Medline: 37817143]
26. Zuckerbrot RA, Cheung A, Jensen PS, Stein REK, Laraque D, GLAD-PC STEERING GROUP. Guidelines for adolescent depression in primary care (GLAD-PC): Part I. Practice preparation, identification, assessment, and initial management. Pediatrics 2018 Mar;141(3):e20174081. [doi: 10.1542/peds.2017-4081] [Medline: 29483200]
27. Hamnvik OPR, Agarwal S, AhnAllen CG, Goldman AL, Reisner SL. Telemedicine and inequities in health care access: the example of transgender health. Transgend Health 2022 Apr;7(2):113-116. [doi: 10.1089/trgh.2020.0122] [Medline: 36644516]
28. Xu C, Miao L, Turner D, DeRubeis R. Urbanicity and depression: a global meta-analysis. J Affect Disord 2023 Nov 1;340:299-311. [doi: 10.1016/j.jad.2023.08.030] [Medline: 37557989]
29. Horvath KJ, Iantaffi A, Swinburne-Romine R, Bockting W. A comparison of mental health, substance use, and sexual risk behaviors between rural and non-rural transgender persons. J Homosex 2014;61(8):1117-1130. [doi: 10.1080/00918369.2014.872502] [Medline: 24380580]
30. Ralston AL, Holt NR, Andrews AR 3rd, et al. Mental health and marginalization stress in transgender and gender diverse adults: differences between urban and non-urban experiences. Psychol Sex Orientat Gend Divers 2024 Mar;11(1):165-176. [doi: 10.1037/sgd0000595] [Medline: 38577413]
31. McGregor K, Rana V, McKenna JL, Williams CR, Vu A, Boskey ER. Understanding family support for transgender youth: impact of support on psychosocial functioning. J Adolesc Health 2024 Aug;75(2):261-266. [doi: 10.1016/j.jadohealth.2024.04.006] [Medline: 38842989]
32. Rickwood D, Thomas K. Conceptual measurement framework for help-seeking for mental health problems. Psychol Res Behav Manag 2012;5:173-183. [doi: 10.2147/PRBM.S38707] [Medline: 23248576]
33. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873. [doi: 10.2196/13873] [Medline: 31742562]

Abbreviations

aPR: adjusted prevalence ratio
BIPOC: Black, Indigenous, and People of Color
PHQ-2: two-item patient health questionnaire
PR: prevalence ratio
TGD: Transgender, nonbinary, gender diverse

Edited by TDA Cardoso; submitted 18.09.24; peer-reviewed by K Clark, ZJG Regencia; revised version received 28.01.25; accepted 24.02.25; published 24.03.25.

Please cite as:
Reisner S, Liu Y, Tham R, Kane K, Cole SW, Boskey ER, Katz-Wise SL, Keuroghlian AS, Xu R
Prevalence and Correlates of Clinically Elevated Depressive Symptoms in a Nationwide Sample of Transgender, Nonbinary, and Gender Diverse Young Adults in the United States: Cross-Sectional Survey Study
Interact J Med Res 2025;14:e66630
URL: https://www.i-jmr.org/2025/1/e66630
doi:10.2196/66630

Original Paper

Citizen Worry and Adherence in Response to Government Restrictions in Switzerland During the COVID-19 Pandemic: Repeated Cross-Sectional Online Surveys

Vanessa Kraege[1,2*], MD, MBA; Céline Dumans-Louis[2,3*], MD; Céline Maglieri[1], MSc; Séverine Bochatay[4]; Marie-Anne Durand[2,5,6], Prof Dr; Antoine Garnier[7,8], MBA, MD; Kevin Selby[2,5], MD; Christian von Plessen[2,5], MD, PhD

[1]Innovation and Clinical Research Directorate, University Hospital of Lausanne, Lausanne, Switzerland

[2]Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

[3]Department of Internal Medicine, Le Groupement Hospitalier de l'Ouest Lémanique, 1260 Nyon, Switzerland

[4]Faculty of Arts, University of Lausanne, Lausanne, Switzerland

[5]Department of Ambulatory Care, Centre universitaire de médecine générale et santé publique, Lausanne, Lausanne, Switzerland

[6]The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Hanover, NH, United States

[7]Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland

[8]Department of Medicine and Specialities, Fribourg Cantonal Hospital, Fribourg, Switzerland

[*]these authors contributed equally

Corresponding Author:
Céline Dumans-Louis, MD
Department of Internal Medicine
Le Groupement Hospitalier de l'Ouest Lémanique
Chemin du Monastier 10
1260 Nyon
Switzerland
Phone: 41 765567810
Email: celine.dumanslouis@ghol.ch

Abstract

Background: Good communication between health authorities and citizens is crucial for adherence to preventive measures during a pandemic. Crisis communication often appeals to worries about negative consequences for oneself or others. While worry can motivate protective behavior, it can also be overwhelming and lead to irrational choices or become a mental health problem. Also, the levels and consequences of worry can differ between different groups of citizens. Little is known about the evolution of worries during the pandemic and adherence to measures in distinct groups.

Objective: This study aimed to evaluate worries in the Swiss population as well as associations between worry levels and citizens' adherence to government restrictions during different phases of the COVID-19 pandemic.

Methods: We carried out an observational study with 4 cross-sectional online surveys of adults in the Canton of Vaud, Switzerland. Questionnaires were distributed through social media and websites during 4 periods: survey 1: April 17 to May 14, 2020; survey 2: May 15 to June 22, 2020; survey 3: October 30 to December 12, 2020; and survey 4: June 18 to December 30, 2021. On visual analog scales from 0 to 100, participants reported worry, self-adherence to pandemic restrictions, and their perceived adherence to others. We used multivariable linear regression, adjusting for age, gender, health literacy, and education to assess associations between self-reported worry, adherence, and study periods.

Results: We collected 7106 responses. After excluding 2377 questionnaires (incomplete, age <18 years, residence outside Vaud), 4729 (66.55%) were analyzed (mean age 47, SD 15.6 years, 63.96% women). Mean worry across the 4 periods was 42/100, significantly higher in women (44.25/100, vs 37.98/100; P<.001) and young people (43.77/100 in those aged 18-39 years, vs 41.69/100; P=.005; in those aged 40-64 years and 39.16/100; P=.002; in those aged >64 years). Worries were higher during survey 1 and survey 3 (52.41/100 and 56.32/100 vs 38.93/100, P<.001; and 35.71/100, P<.001) than during survey 2 and survey 4, respectively. This corresponds to pandemic peaks during which federal restrictions were better followed with self-reported adherence of 84.80/100 and 89.59/100 in survey 1 and survey 3 versus 78.69/100 (P<.001) and 78.64/100 (P<.001) in survey 2

and survey 4. A 2.9-point increase in worry score, adjusted for the pandemic period, gender, age, education, and health literacy, was associated with a 10-point increase in personal adherence score (95% CI 2.5-3.2; *P*<.001).

Conclusions: Worries were higher in women, young people, and during the peak of the COVID-19 pandemic. Higher worry levels were associated with increased self-reported adherence to federal restrictions. Authorities should consider population worry levels and population subgroups in the planning and design of pandemic communication.

(Interact J Med Res 2025;14:e55636) doi:10.2196/55636

KEYWORDS

COVID-19 pandemic; citizens; worry; anxiety; communication; prevention; adherence; restrictions; Switzerland; cross sectional; online survey; survey; Swiss; adults; questionnaire; social media; linear regression; age; gender; health literacy; education; women; young people

Introduction

Effective communication between health authorities and the population is crucial to achieving public health goals during a pandemic. Providing clear, consistent, and reliable information that motivated behavior changes without triggering resistance was a major challenge during the COVID-19 pandemic. Sanitary restrictions were often rapidly issued and modified to contain the spread of the disease [1-3]. Citizens were expected to make drastic behavioral changes.

Public health authorities stressed the seriousness and risks of the pandemic to justify restrictions and encourage citizens' adherence to them. In support of such an approach, the Health Belief Model argues that preventive health behaviors are influenced by perceived susceptibility to illness, the severity of the disease, benefits of and barriers to health-promoting actions, cues to action, as well as self-efficacy [4]. Also, during the COVID-19 pandemic, protective behaviors were associated with these factors, especially when the "perceived benefit" of a measure was clear [5].

Appealing to emotions such as fear can hence be a persuasive way of motivating respect for protective measures [6,7]. Indeed, fear leads to behavioral change if people feel capable of dealing with the threat, while they become defensive when feeling helpless and incapable of acting [8-10]. Overdriven or ill-conceived fear-based communication may even provoke counterproductive behavior.

Levels of anxiety, worry, and stress were high during the pandemic. According to a systematic review and meta-analysis, anxiety prevalence was around 30% worldwide after the first COVID-19 wave [11]. Others have confirmed these findings [12,13]. Among professionally active persons, 42% of participants reported being worried about the COVID-19 pandemic in August and September 2020 [14]. Young adults in the city of Zurich, Switzerland, reported elevated stress levels in April 2020, in the aftermath of the first wave [15], as was found more generally in Swiss adults too [16]. In terms of risk factors, anxiety was higher in women, younger people, and vulnerable persons [17-19].

In our previously published cross-sectional population survey, performed during the first wave of the pandemic, we found high self-reported adherence to official restrictions, which increased with age and level of worry [20]. As in the aforementioned studies, worry was high, particularly among people in isolation and with lower health literacy. Nearly half of the respondents felt that government responses were adequate or, associated with higher levels of worry, even insufficient. Neither the aforementioned nor our cross-sectional study could determine the evolution of these associations throughout the pandemic.

Thus, we conducted surveys during different phases of the pandemic to describe the evolution of worries in the Swiss population as well as associations between worries and adherence to governmental restrictions. Our overall aim was to contribute new insights to this understudied area to help improve crisis communication during future pandemics.

Methods

Study Design and Setting

We conducted repeated cross-sectional online surveys in Vaud, a French-speaking canton of 823,000 inhabitants (2021) in Switzerland. We launched 4 surveys between April 2020 and December 2021: survey 1 between April 17 and May 14, 2020 (4 weeks); survey 2 between May 15 and June 22, 2020 (5.5 weeks); survey 3 between October 30 and December 1, 2020 (4.5 weeks); and survey 4 between June 18 and December 30, 2021 (28 weeks, Table 1). Some of the survey items were adapted or replaced to capture changes in federal measures. We followed the CHERRIES (Checklist for Reporting Results of Internet e-Surveys) guidelines [21]. Self-reported worry was an outcome of the study of worry levels and an exposure variable for the study of associations between worry and self-reported adherence during these 4 COVID-19 pandemic periods.

Table 1. Surveyed periods and sentinel pandemic-related events.

Survey	Period	Number of weeks	Sentinel events
Survey 1: End of the first pandemic wave	April 17, 2020, to May 14, 2020	4	• March 16, 2020: Semiconfinement, only essential shops open, gatherings of a maximum of 5 people • April 27, 2020: Partial reopening of shops • May 11, 2020: Reopening of schools
Survey 2: After the first pandemic wave	May 15, 2020, to June 22, 2020	5.5	• June 19, 2020: End of an extraordinary situation
Survey 3: During the second pandemic wave	October 30, 2020, to December 1, 2020	4.5	• Mandatory wearing of masks in indoor public spaces; gatherings limited to 15 people
Survey 4: Following pandemic waves	June 18, 2021, to December 30, 2021	28	• Vaccination available to all, use of COVID-19 vaccination certificate

Table 1 presents the COVID-19 waves and their duration in the French-speaking part of Switzerland. The first and second forms were distributed at the end of the first wave, corresponding to the gradual emergence from confinement. The third form was distributed over a longer period, which included the second wave and the resumption of restrictive measures. The fourth form was distributed once vaccination was available for the entire population.

Participant Recruitment

Using a weblink, we distributed the surveys on the social media platforms of multiple community organizations to collect a convenience sample of the population. These organizations were a regional consumer organization, regional disease leagues for cancer and diabetes, the association of senior citizens as well as the cantonal websites for the COVID-19 testing and vaccination decisions. These cantonal sites were used by large portions of the population. The organizations advertised the study through links on their websites and some social media accounts. The links were accompanied by a short explanation of the study and its purpose. No incentives to participate were given. The online interface for the survey was created in REDCap (Research Electronic Data Capture, Vanderbilt University).

Surveys

The development and testing of the survey are described in our previous publication with results from survey 1 [20]. Each survey was submitted to 5 nonmedical persons to test the understandability of questions. There was no review step for this short questionnaire. The first, second, and fourth surveys had 20 items, and the third survey had 25 items. A total of 12 items remained unchanged throughout all surveys. We included demographic data, such as age, sex, number of persons per household, canton of residence, level of education, literacy, and whether the respondent had been tested for the COVID-19 pandemic. For the literacy question, we used a validated item from Chew et al [22]. Employment status was included in surveys 2-4. Respondents rated (1) worry about the pandemic situation, (2) self-reported adherence to government restrictions, and (3) perceived adherence of others to government restrictions, on visual analog scales from 0 to 100 (0=not at all; 100=in all situations). Items were not randomized. Participants could go back to earlier questions at any time. Adaptive questioning was used for several items. The first 2 surveys took up 6 screens, and the 2 latter, 7. REDCap automatically generated a completeness variable if participants went all the way to the end of the survey. Anyone who opened the survey generated a response. We did not determine unique site visitors, establish view or participation rates, or IP address checks, as surveys were entirely anonymous and the risk of repeating them was low. The time used to fill them in was not registered. No cookies were used, and there were no other techniques to analyze the log file of our database. Statistical corrections were not used. The 4 surveys can be found in Multimedia Appendices 1-4.

Statistical Analyses

The item “Level of education” was dichotomized into “university or college education” or other. Health literacy was dichotomized according to ease of answering a medical form on one’s own: low literacy (“never,” “rarely,” or “sometimes” at ease) and “high literacy” (“often” or “always”). This was based on a validated item [22] and our previous article [20].

We limited our analysis to complete questionnaires for 2 reasons. First, the survey was distributed using an online link on government websites and a large number of persons clicked the link but only completed 1 or 2 questions. Second, no single question had many missing responses and we preferred to maintain consistency across surveys. We calculated means with SD and frequencies with IQR as appropriate. Independence between surveys was tested with the *t* test for continuous variables (eg, age), and with the chi-square test for gender, education, and health literacy. We performed linear regressions to analyze associations between the 4 periods and level of worry (model A), self-reported adherence (model B), and perceived adherence of others to restrictions (model C). First, we performed univariate linear regression, followed by multivariable linear regression controlling for age (grouped as 18-39 years, 40-64 years, and 65 years or older), sex (dichotomized male-female), level of education (dichotomized university education or other), and health literacy (dichotomized high or low health literacy). We subsequently used the margins command in Stata (StataCorp) to report absolute differences in the predicted levels of worry or adherence with each model. The level of significance was set to $P<.05$. Statistical analyses were performed in Microsoft Excel and Stata (version 16.1).

Ethical Considerations

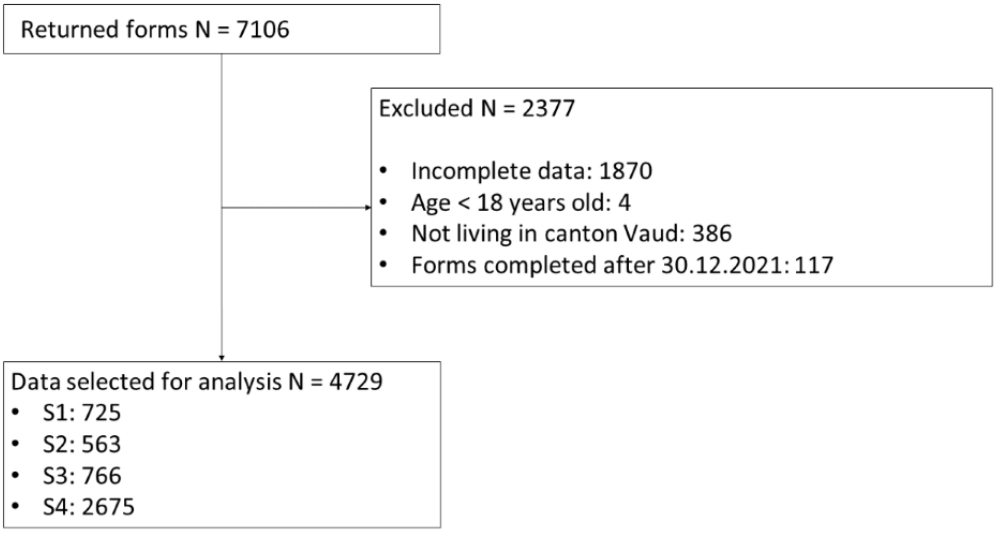
According to the Cantonal Commission on Ethics in Research Involving Human Beings of the Canton of Vaud, Switzerland the study was exempted from ethical review because it did not qualify as human subjects research and all data collection was anonymous (2024-010901).

Results

Participant Characteristics

Citizens completed 7106 surveys between April 17, 2020, and December 20, 2021. After the exclusion of minors, persons living outside the Canton, and incomplete questionnaires, 4729 (66%) surveys remained. The number of questionnaires per period ranged between 563 and 2675 (Figure 1).

Figure 1. Study participants.



The participants were more often women (3025/4729, 63.96%) and between 40 and 64 years old (2442/4729, 51.64%). Furthermore, 2526/4729 (53.42%) of participants had attended university or college, and 4510/4729 (95.37%) reported high health literacy. The participants were younger in survey 3 (mean age 43.2, SD 14.6 years; $P<.001$) than in the other surveys (survey 1: mean age 47.9, SD 14.6 years; survey 2: mean age 47.3, SD 15.6 years; and survey 4: mean age 48.7, SD 15.3 years; Table 2).

Table 2. Demographics, education, health literacy, and employment of participants (n=4729).

Variables	All surveys	Survey 1[a]	Survey 2[b]	Survey 3[c]	Survey 4[d]
Age (years), n (%)					
18-39	1574 (33.3)	244 (33.7)	202 (35.9)	325 (42.4)	803 (30)
40-64	2442 (51.6)	365 (50.3)	268 (47.6)	369 (48.2)	1440 (53.8)
≥65	713 (15.1)	116 (16)	93 (16.5)	72 (9.4)	432 (16.1)
Years, mean (SD)	47.5 (15.3)	47.9 (14.6)	47.3 (15.6)	43.2 (14.6)	48.7 (15.3)
Gender, n (%)					
Male	1698 (35.9)	168 (23.2)	178 (31.6)	264 (34.5)	1088 (40.7)
Female	3025 (64)	557 (76.8)	384 (68.2)	502 (65.5)	1582 (59.1)
Other[e]	6 (0.1)	0 (0)	1 (0.2)	0 (0)	5 (0.2)
Education, n (%)					
Obligatory school or less	256 (5.4)	12 (1.7)	21 (3.7)	28 (3.7)	195 (7.3)
Apprenticeship	1379 (29.2)	171 (23.6)	131 (23.3)	209 (27.3)	868 (32.4)
High-school graduation	502 (10.6)	69 (9.5)	47 (8.3)	85 (11.1)	301 (11.3)
University or college	2526 (53.4)	466 (64.3)	358 (63.6)	434 (56.7)	1268 (47.4)
I do not know	66 (1.4)	7 (1)	6 (1.1)	10 (1.3)	42 (1.6)
Health literacy, n (%)					
Low health literacy	219 (4.6)	52 (7.2)	33 (5.9)	74 (9.7)	344 (12.9)
High health literacy	4510 (95.4)	673 (92.8)	529 (94.1)	692 (90.3)	2326 (87.1)
Employment[f], n (%)					
Full-time work	1707 (42.6)	—[g]	225 (40.0)	347 (45.3)	1135 (42.4)
Part-time work	674 (16.8)	—	102 (18.1)	136 (17.8)	436 (16.3)
Housewife and husband	125 (3.1)	—	18 (3.2)	24 (3.1)	83 (3.1)
Self-employed	303 (7.6)	—	37 (6.6)	46 (6.0)	220 (8.2)
Student	241 (6)	—	31 (5.5)	72 (9.4)	138 (5.2)
Employment status, n (%)					
Unemployed and currently looking for a job	133 (3.3)	—	23 (4.1)	19 (2.5)	91 (3.4)
Unemployed and not currently seeking employment	57 (1.4)	—	8 (1.4)	14 (1.8)	35 (1.3)
Incapacity	129 (3.2)	—	11 (2)	26 (3.4)	92 (3.4)
Retired	619 (15.5)	—	105 (18.7)	81 (10.6)	433 (16.2)
Unknown	15 (0.4)	—	3 (0.5)	1 (0.1)	11 (0.4)

[a]Survey 1: April 17 to May 14, 2020.

[b]Survey 2: May 15 to June 22, 2020.

[c]Survey 3: October 30 to December 1, 2020.

[d]Survey 4: June 18, 2021, to December 30, 2021.

[e]These were excluded from the regression analyses.

[f]Employment data were not collected during the first survey.

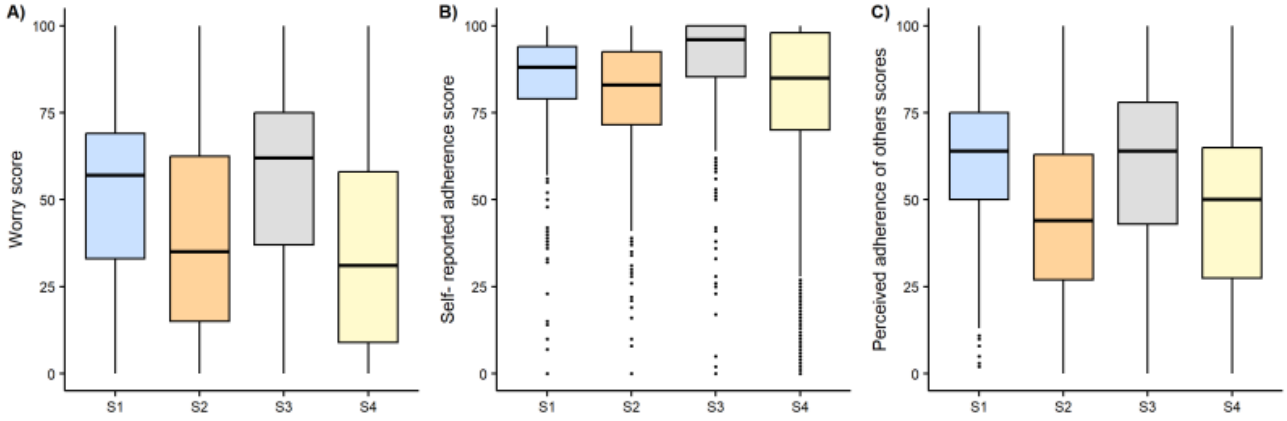
[g]Not available.

Main Results

Self-Reported Worry

Overall, the mean level of self-reported worry was 42.0% (SD 28.9). Upon univariate regression, self-reported worry levels differed significantly across surveys, with significantly higher levels in survey 1 (52%, 95% CI 50-54) and survey 3 (56%, 95% CI 54-58; Figure 2). Upon multivariable regression, the female gender was associated with a 4-point increase in level of worry (95% CI 2-5 points). Worry levels were 2 (95% CI 1-4) and 4 (95% CI 1-6) points lower among respondents aged 40 to 64 and over 64, respectively, when compared with the 18- to 39-year group. Higher health literacy was associated with a 3-point lower worry level (95% CI –6 to –1). Education was not associated with significant changes in self-reported level of worry (Table S5 in Multimedia Appendix 5).

Figure 2. Boxplots on a 0-100 scale of (A) worry, (B) self-reported adherence, and (C) perceived adherence of others to restrictions during four COVID-19 pandemic periods (April 17, 2020, to December 30, 2021), Switzerland. S: survey; survey 1: April 17 to May 14, 2020; survey 2: May 15 to June 22, 2020; survey 3: October 30 to December 1, 2020; survey 4: June 18, 2021, to December 30, 2021.



Self-Reported Adherence to Restrictions

Overall, respondents evaluated their own adherence to government restrictions at 81.4% (SD 21.1). Self-reported adherence was significantly (P<.001) higher in survey 1 (mean 84.8%, SD 14.2) and survey 3 (mean 89.6%, SD 15.5; Figure 2). A 2.9-point increase in the worry score was associated with a 10-point increase in self-reported adherence (95% CI 2.5-3.2; P<.001) after adjusting for the pandemic period, gender, age, education, and health literacy. These effects were more pronounced in women and older participants (Table S5 in Multimedia Appendix 5). Moreover, both older age categories (40-64 years and more than 64 years) were associated with a 7-point higher self-reported adherence than in the 18-39 years group (95% CI 6-8 for 40- to 64-year-olds and 6-9 for >64-year-olds). Higher health literacy was associated with a 4-point increase in self-reported adherence (95% CI 2-6) while the educational level was not (Table S5 in Multimedia Appendix 5).

Perceived Adherence of Others to Restrictions

Overall, respondents evaluated the adherence of others to government restrictions at 50.4% (SD 24.5%). Evaluated adherence did not differ between survey 1 and survey 3 nor between survey 2 and survey 4 but was significantly higher (P<.001) in survey 1 (mean 60.1%, SD 20.0%) and survey 3 (59.6%, SD 24.6%) than in survey 2 (45.1%, SD 22.5%) and survey 4 (46.2%, SD 24.5%; Figure 2). Adjusting for participant characteristics, age groups 40-64 years and >64 years were associated with, respectively, a 5- or 7-point higher perceived adherence of others (95% CI 3-6 and 5-9, respectively). A 10-point increase in worry level was associated with a 1-point decrease in perceived adherence of others to restrictions (95% CI –0.9 to –0.4). Higher education level was associated with a 1.5-point higher perceived adherence of others (95% CI, 0-3), whereas gender and health literacy were not (Table S5 in Multimedia Appendix 5).

Changes to the Daily Lives of Respondents

During the first period (survey 1), most respondents had experienced changes in their daily life (Table S6 in Multimedia Appendix 6). Later (surveys 2-4), 35% to 49% had either lost their job or had to close their business and 21% to 34% had lost part of their income. Respondents also reported feeling isolated, lonelier, and less productive during surveys 2 to 4. In decreasing order of importance, concerns during survey 3 were for “vulnerable people,” “living conditions,” the “economy,” “self and family,” “working conditions,” and the “possibility of another wave.” During survey 4, these concerns were similar but generally rated lower. Interestingly, “deterioration of working conditions” moved up from fifth to third rank, which had been “self and family” in survey 3 (Multimedia Appendix Table S7 in Multimedia Appendix 7 and Table S8 in Multimedia Appendix 8).

Discussion

Principal Findings

We conducted online surveys during different phases of the COVID-19 pandemic to describe the evolution of worry levels and to assess how these were associated with adherence to government restrictions. In Switzerland, the self-reported worry was highest during the first and second pandemic waves, corresponding to survey 1 and survey 3, at times of many COVID-19–related hospitalizations and deaths, and when a vaccine was not yet available. Women and younger people

reported higher levels of worry than men and older people. Education did not influence worry levels, while lower health literacy was associated with higher worry. Higher worry levels were associated with higher self-reported and perceived adherence of others to federal restrictions.

We found elevated worry levels during the more dramatic phases of the pandemic. In a systematic review covering 204 countries in 2020, higher anxiety was associated with higher COVID-19 incidence [23], a finding that was confirmed by Salanti et al [13]. In March 2020, Fitzpatrick et al [24] found that in a national sample in the United States, worry was higher in the regions with high COVID-19 incidence. In Ontario, Canada, COVID-19–related worry in young persons also increased during the early phases of the pandemic and then again in the autumn of 2020, when the incidence was higher [25]. So, anxiety and worry varied during the pandemic and increased repeatedly with the rising incidence of COVID-19. This is in line with findings on decreased mental health on a larger scale during the pandemic [11,26].

In periods of increased worry, we found higher self-reported adherence to government restrictions. Similarly, a study in Saudi Arabia described an association between higher anxiety levels and preventive practices among health care workers [27]. Another study identified fear as a predictor of behavioral change [28]. The association between worry and adherence in our study could indicate that worry was not overwhelming and that citizens felt in control of risks by respecting restrictions. We cannot exclude that this might have been different with higher anxiety levels.

Considering population subgroups, young adults were often the most anxious despite being less at risk of hospitalization or death [17,18,23,29]. Young people were worried about social isolation and develop depressive symptoms during school closings [30]. In Switzerland, students were concerned about whether they would be able to finish the 2020 university year [31], and lockdowns as well as their socioeconomic consequences were stressful for students [15]. Apart from concerns about the future, young people were not only worried about their own health but also about that of relatives. For example, in a study in Zurich, Switzerland, students were more concerned about the health of their parents and grandparents than their own [32]. Also, our finding of higher worry levels in women echoes several publications [17-19,33]. General factors potentially contributing to worry were the increasing risk of unemployment or loss of income, as well as loneliness and feeling less productive. One-third of respondents had lost part of their income by surveys 2 and 3 (31%, and 34%, respectively), with a slightly better situation in survey 4 (21%). The reported feelings of isolation, loneliness, and being less productive could further contribute to worry in some respondents. For example, studying for exams through online classes only, without any campus life, can be a source of worry compared with when stress from studying and exams is compensated by in-person interactions with teachers and colleagues. Four years after the pandemic, Sayed et al. [34] insist on the importance of addressing mental health of children and young adults during global crises and of recognizing long-term impacts. They further emphasize the need for research and public health prioritization of these important topics.

Overall, our findings are in line with publications that highlight the importance of addressing the many individual and collective aspects that influenced mental health during the COVID-19 pandemic, such as isolation, loneliness, and fear [35]. Many individuals demonstrated remarkable resilience, allowing society to avoid a general increase in loneliness [36]. However, population estimates may mask individual heterogeneity; loneliness is indeed a major public health concern and must be considered as a negative determinant of health [35]. Even though the pandemic is over, we must not forget its long-term effects on mental health and public health authorities should consider the differing impact of governmental decisions on the general population versus on individuals with pre-existing mental health conditions [37]. Our results of the worries of citizens and adherence to pandemic measures can be useful in preparing for future pandemics, for example, in considering criteria for and potential impact of restrictions on different subgroups of the population.

Strengths and Limitations

The strengths of our study were repeated surveys with similar questions and our relatively large sample size, allowing us to examine subgroups of the population.

Concerning limitations, data were collected through the online distribution of surveys in a simple, cost-effective, and feasible way during the rapidly evolving pandemic. Participation was more attractive to women and younger people with higher literacy and education. Also, participation was variable during data collection periods. The participant sample was more representative in the last and longest period (survey 4). This selection and variation need to be considered in the interpretation of our data which are prone to desirability, information, and selection bias. Different distribution channels and methods are probably needed for disadvantaged populations, as we showed in a recent study using our survey in a population of refugees and migrants [38]. Finally, another inherent limitation of our anonymous data collection is that we could not follow a cohort of persons throughout the pandemic.

Conclusion

Worry reached moderate levels and varied with COVID-19 incidence during the pandemic. Higher worry levels were associated with increased self-reported and perceived adherence of others to government restrictions. Younger people and women reported higher worry levels. Authorities should take population worry levels into account in planning and designing pandemic communication. Adapting communication to population subgroups should be considered for future health crises.

Acknowledgments

We would like to thank our community partners for distributing our electronic survey, specifically Yannis Papa Daniel from the Fédération Romande des Consommateurs and Dr Alix Miauton for the Coronacheck and Coronavax websites. We would also like to thank everyone who answered the surveys.

There was no specific financial support for the completion of this study.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CDL and VK wrote most of the manuscript. CM did most of the statistical analyses and wrote part of the article. SB contributed to analyzing and interpreting data and did part of the statistical analyses. M-AD contributed to designing the study and the questionnaires, collected data, and edited the manuscript. AG contributed to the logistics of the research team, interpreted the data, and edited the manuscript. KS contributed to designing the study and the questionnaires, collected data, and did statistical analyses. CVP was the instigator of the study and contributed to the study and questionnaire design, data collection, and manuscript writing. All authors revised the manuscript for intellectual content and approved it before submission. The authors indicate that they did not use generative artificial intelligence when writing their manuscript.

Conflicts of Interest

M-AD has contributed to the development of Option Grid patient decision aids and receives consulting income from EBSCO Health and royalties. KS receives salary support from the Leenaards Foundation. EBSCO Information Services sells subscription access to Option Grid patient decision aids.

Multimedia Appendix 1
Survey 1.
[PDF File (Adobe PDF File), 362 KB - ijmr_v14i1e55636_app1.pdf]

Multimedia Appendix 2
Survey 2.
[PDF File (Adobe PDF File), 343 KB - ijmr_v14i1e55636_app2.pdf]

Multimedia Appendix 3
Survey 3.
[PDF File (Adobe PDF File), 384 KB - ijmr_v14i1e55636_app3.pdf]

Multimedia Appendix 4
Survey 4.
[PDF File (Adobe PDF File), 370 KB - ijmr_v14i1e55636_app4.pdf]

Multimedia Appendix 5
Linear regression results for three models; all on visual analogue scales from 0 to 100 (0=not at all; 100=in all situations).
[PDF File (Adobe PDF File), 77 KB - ijmr_v14i1e55636_app5.pdf]

Multimedia Appendix 6
Perceived changes of daily life during the first pandemic wave.
[PDF File (Adobe PDF File), 45 KB - ijmr_v14i1e55636_app6.pdf]

Multimedia Appendix 7
Impact of restrictions on daily life, S2, S3, S4.
[PDF File (Adobe PDF File), 40 KB - ijmr_v14i1e55636_app7.pdf]

Multimedia Appendix 8
Concerns of respondents, S3, S4.
[PDF File (Adobe PDF File), 40 KB - ijmr_v14i1e55636_app8.pdf]

References

1. WHO. Listings of WHO's response to COVID-19. URL: https://www.who.int/news/item/29-06-2020-covidtimeline [accessed 2023-12-11]
2. Recommendations to Member States to improve hand hygiene practices to help prevent the transmission of the COVID-19 virus: interim guidance, 1 April 2020. Geneva, Switzerland: World Health Organization; 2020. URL: https://iris.who.int/handle/10665/331661
3. World Health Organization. Considerations for quarantine of contacts of COVID-19 cases: interim guidance, 25 June 2021. Geneva, Switzerland: World Health Organization; 2021. URL: https://iris.who.int/handle/10665/342004
4. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
5. Zewdie A, Mose A, Sahle T, Bedewi J, Gashu M, Kebede N, et al. The health belief model's ability to predict COVID-19 preventive behavior: a systematic review. SAGE Open Med 2022;10::20503121221113668 [FREE Full text] [doi: 10.1177/20503121221113668] [Medline: 35898953]
6. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
7. Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, et al. Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. Psychol Bull 2015;141(6):1178-1204 [FREE Full text] [doi: 10.1037/a0039729] [Medline: 26501228]
8. Bigsby E, Albarracín D. Self- and response efficacy information in fear appeals: a meta-analysis. J Commun 2022;72(2):241-263. [doi: 10.1093/joc/jqab048]
9. Lunn PD, Belton CA, Lavin C, McGowan FP, Timmons S, Robertson DA. Using behavioral science to help fight the coronavirus. JBPA 2020;3(1) [FREE Full text] [doi: 10.30636/jbpa.31.147]
10. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
11. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
12. Santabárbara J, Lasheras I, Lipnicki DM, Bueno-Notivol J, Pérez-Moreno M, López-Antón R, et al. Prevalence of anxiety in the COVID-19 pandemic: an updated meta-analysis of community-based studies. Prog Neuropsychopharmacol Biol Psychiatry 2021;109:110207 [FREE Full text] [doi: 10.1016/j.pnpbp.2020.110207] [Medline: 33338558]
13. Salanti G, Peter N, Tonia T, Holloway A, White IR, Darwish L, MHCOVID Crowd Investigators, MHCOVID Crowd Investigators†. The impact of the COVID-19 pandemic and associated control measures on the mental health of the general population : a systematic review and dose-response meta-analysis. Ann Intern Med 2022;175(11):1560-1571 [FREE Full text] [doi: 10.7326/M22-1507] [Medline: 36252247]
14. Yong X, Zhang L. COVID-19 worry and mental health among the economically active population in Guangdong, China. Front Public Health 2022;10:882177 [FREE Full text] [doi: 10.3389/fpubh.2022.882177] [Medline: 35586016]
15. Shanahan L, Steinhoff A, Bechtiger L, Murray AL, Nivette A, Hepp U, et al. Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study. Psychol Med 2022;52(5):824-833 [FREE Full text] [doi: 10.1017/S003329172000241X] [Medline: 32571438]
16. De Quervain DJF, Aerni A, Amini E, Bentz D, Coynel D, Gerhards C, et al. The Swiss Corona Stress Study Internet. Open Science Framework 2020 [FREE Full text] [doi: 10.31219/osf.io/jqw6a]
17. Moghanibashi-Mansourieh A. Assessing the anxiety level of Iranian general population during COVID-19 outbreak. Asian J Psychiatr 2020;51:102076 [FREE Full text] [doi: 10.1016/j.ajp.2020.102076] [Medline: 32334409]
18. Mazza C, Ricci E, Biondi S, Colasanti M, Ferracuti S, Napoli C, et al. A nationwide survey of psychological distress among Italian people during the COVID-19 pandemic: immediate psychological responses and associated factors. Int J Environ Res Public Health 2020;17(9):3165 [FREE Full text] [doi: 10.3390/ijerph17093165] [Medline: 32370116]
19. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
20. Selby K, Durand MA, Gouveia A, Bosisio F, Barazzetti G, Hostettler M, et al. Citizen responses to government restrictions in Switzerland during the COVID-19 pandemic: cross-sectional survey. JMIR Form Res 2020;4(12):e20871 [FREE Full text] [doi: 10.2196/20871] [Medline: 33156809]
21. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
22. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]

23. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
24. Fitzpatrick KM, Drawve G, Harris C. Facing new fears during the COVID-19 pandemic: the state of America's mental health. J Anxiety Disord 2020;75:102291 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102291] [Medline: 32827869]
25. Hawke LD, Szatmari P, Cleverley K, Courtney D, Cheung A, Voineskos AN, et al. Youth in a pandemic: a longitudinal examination of youth mental health and substance use concerns during COVID-19. BMJ Open 2021;11(10):e049209 [FREE Full text] [doi: 10.1136/bmjopen-2021-049209] [Medline: 34716160]
26. Mental Health and COVID-19: early evidence of the pandemic's impact: Scientific brief, 2 March 2022. Geneva, Switzerland: World Health Organization; 2022. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 [accessed 2024-12-27]
27. Alshagrawi S, Alhowti S. Knowledge, anxiety, and practice during the COVID-19 crisis among HCWs in Saudi Arabia. J Educ Health Promot 2022;11:384 [FREE Full text] [doi: 10.4103/jehp.jehp_308_22] [Medline: 36618461]
28. Harper CA, Satchell LP, Fido D, Latzman RD. Functional fear predicts public health compliance in the COVID-19 pandemic. Int J Ment Health Addict 2021;19(5):1875-1888 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
29. Hyland P, Shevlin M, McBride O, Murphy J, Karatzias T, Bentall RP, et al. Anxiety and depression in the republic of Ireland during the COVID-19 pandemic. Acta Psychiatr Scand 2020;142(3):249-256. [doi: 10.1111/acps.13219] [Medline: 32716520]
30. Hawes MT, Szenczy AK, Klein DN, Hajcak G, Nelson BD. Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. Psychol Med 2022;52(14):3222-3230 [FREE Full text] [doi: 10.1017/S0033291720005358] [Medline: 33436120]
31. Amendola S, von Wyl A, Volken T, Zysset A, Huber M, Dratva J. A longitudinal study on generalized anxiety among university students during the first wave of the COVID-19 pandemic in Switzerland. Front Psychol 2021;12:643171 [FREE Full text] [doi: 10.3389/fpsyg.2021.643171] [Medline: 33776867]
32. Dratva J, Zysset A, Schlatter N, von Wyl A, Huber M, Volken T. Swiss university students' risk perception and general anxiety during the COVID-19 pandemic. Int J Environ Res Public Health 2020;17(20):7433 [FREE Full text] [doi: 10.3390/ijerph17207433] [Medline: 33066008]
33. Robinson E, Sutin AR, Daly M, Jones A. A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. J Affect Disord 2022;296:567-576 [FREE Full text] [doi: 10.1016/j.jad.2021.09.098] [Medline: 34600966]
34. Sayed AA, El-Gendy AA, Aljohani AK, Haddad RA, Taher OH, Senan AM. The effects of COVID-19 on the mental health of children and adolescents: a review. Cureus. 2024. URL: https://www.cureus.com/articles/207363-the-effects-of-covid-19-on-the-mental-health-of-children-and-adolescents-a-review [accessed 2024-06-13]
35. Lederman Z. Loneliness at the age of COVID-19. J Med Ethics 2023;49(9):649-654. [doi: 10.1136/jme-2022-108540] [Medline: 36600628]
36. Luchetti M, Lee JH, Aschwanden D, Sesker A, Strickhouser JE, Terracciano A, et al. The trajectory of loneliness in response to COVID-19. Am Psychol 2020;75(7):897-908 [FREE Full text] [doi: 10.1037/amp0000690] [Medline: 32567879]
37. Maher A. Tracing the bounds of distress: mental health and the Lancet commission on lessons for the future from the COVID-19 pandemic. J Glob Health 2024;14:03022 [FREE Full text] [doi: 10.7189/jogh.14.03022] [Medline: 38751302]
38. Morisod K, Durand MA, Selby K, Le Pogam MA, Grazioli VS, Sanchis Zozaya J, et al. Asylum seekers' responses to government COVID-19 recommendations: a cross-sectional survey in a Swiss Canton. J Immigr Minor Health 2023;25(3):570-579 [FREE Full text] [doi: 10.1007/s10903-022-01436-3] [Medline: 36508030]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet e-Surveys
REDCap: Research Electronic Data Capture

Edited by T de Azevedo Cardoso; submitted 22.12.23; peer-reviewed by S Haile, JV Sánchez-Ortí; comments to author 23.05.24; revised version received 13.06.24; accepted 25.09.24; published 07.01.25.

Please cite as:
Kraege V, Dumans-Louis C, Maglieri C, Bochatay S, Durand MA, Garnier A, Selby K, von Plessen C
Citizen Worry and Adherence in Response to Government Restrictions in Switzerland During the COVID-19 Pandemic: Repeated Cross-Sectional Online Surveys
Interact J Med Res 2025;14:e55636
URL: https://www.i-jmr.org/2025/1/e55636
doi:10.2196/55636
PMID:

Original Paper

French Versions of 4 English Questionnaires on Problematic Smartphone Use: Cross-Cultural Linguistic Translation and Adaptation Study

Islam El Boudi[1], MEng; Mathilde Riant[1], PhD; Alexandre Bellier[1,2], MD, PhD; Nicolas Vuillerme[1,3], PhD

[1]Faculty of Medicine, AGEIS, Grenoble Alpes University, Grenoble, France

[2]Inserm CIC 1406, Grenoble Alpes University Hospital (CHUGA), Grenoble Alpes University, 38000 Grenoble, France

[3]Institut Universitaire de France, 75005 Paris, France

Corresponding Author:
Islam El Boudi, MEng
Faculty of Medicine
AGEIS, Grenoble Alpes University
Jean Roget Bldg, 3rd Fl.
Grenoble, 38000
France
Phone: 33 695685616
Email: islam.el-boudi@univ-grenoble-alpes.fr

Abstract

Background: Excessive use of smartphones is recognized as a major problem in our modern society and can have dramatic consequences on the health of adolescents and young adults. Measuring problematic smartphone use in research and clinical practice is generally operationalized with self-reported questionnaires. In order to comprehensively assess the issue of problematic smartphone usage within the French population, it is imperative to employ validated French-language questionnaires. However, at this point, existing questionnaires are primarily available in English. Furthermore, to the best of our knowledge, these English questionnaires have yet to undergo validation processes for French-speaking cohorts.

Objective: The aim of this study was to perform a cross-cultural translation of the Smartphone Addiction Scale, Nomophobia Questionnaire, Problematic Use of Mobile Phones scale, and Smartphone Addiction Proneness Scale to French.

Methods: The translation process was performed using the forward/backward method. The first translation phase involved asking 4 independent French translators to translate the original English version of the questionnaires into French. In the second phase, the French version was backtranslated to English by a native English speaker. In the third phase, 2 concept experts were asked to comment and suggest modifications to the statements if necessary. Finally, the last version of the translated questionnaires was presented to 18 participants to assess the clarity, intelligibility, and acceptability of the translations.

Results: During the forward translation step, the translation differences were minor. During the backward translation, the English native speaker correctly backtranslated 18 of the 33 items of the Smartphone Addiction Scale, 17 of the 20 items of the Problematic Use of Mobile Phones scale, and 13 of the 15 items of the Smartphone Addiction Proneness Scale. Backtranslation for the Nomophobia Questionnaire was less satisfactory, with only 10 out of 20 items that were correctly backtranslated. The linguistic verification step revealed a minimal modification for the 4 questionnaires. The participants also suggested few improvements that we have considered for the final version. We produced the final version directly after this step.

Conclusions: We successfully adapted and effectively translated 4 questionnaires that assess problematic smartphone use to French. This step is a prerequisite for the validation of the French questionnaires. These adapted measures can serve as valuable research instruments for investigating and addressing issues related to problematic smartphone use in French-speaking countries and for making international comparisons.

(Interact J Med Res 2025;14:e53958) doi:10.2196/53958

KEYWORDS

problematic use; smartphone; French questionnaire; linguistic translation; forward/backward process; mobile phone

Introduction

Today, the excessive use of smartphones and screens, in general, is recognized as a major problem in our society. Recent world events that have led to individuals being confined to their homes for several months have intensified the amount of scientific and political concerns surrounding this issue. Excessive smartphone use can have dramatic consequences for the health and cognitive development of adolescents and young adults [1]. Indeed, difficulty in regulating smartphone use can lead to serious pathological disorders such as addiction to video games [2].

Although health care professionals, researchers, and politicians recognize that excessive smartphone use is problematic, its legitimacy as a pathological disorder is not yet recognized by the American Psychiatric Association [2]. For this reason, the notion of problematic smartphone use is favored by a large number of researchers. Problematic smartphone use does, however, refer to symptoms of dependence such as loss of control, overuse, increased tolerance (ie, spending more time on the smartphone to be satiated), or withdrawal symptoms once individuals are no longer in possession of their smartphones [3].

In order to identify individuals at risk of problematic use, the measurement of smartphone use practices is generally operationalized in research by using self-reported questionnaires. Three very recently published systematic reviews [3-5] identified 4 measurement scales most commonly used to measure problematic smartphone use, 3 of which are used with young adult students, that is, the Smartphone Addiction Scale [6], the Nomophobia Questionnaire [7], and the Problematic Use of Mobile Phones scale [8], and 1 for adolescents, that is, the Smartphone Addiction Proneness Scale [9]. What these 4 scales have in common is that they assess problematic smartphone use across several dimensions, often inspired by the fourth and fifth versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV and DSM-V) [2,10].

The Smartphone Addiction Scale was designed by Kwon et al [6] based on the diagnostic criteria for internet gaming addiction in DSM-IV [10]. It measures problematic smartphone use along 6 dimensions: disruption of daily life due to smartphone use, positive anticipation of use, withdrawal symptoms, overuse, tolerance, and importance of the virtual world for social relationships. This scale has the advantage of being validated in several languages, including English [6], Arabic [11], Brazilian [12], Malaysian [13], and Persian [14]. It also has good cross-cultural consistency, with Cronbach α ranging from 0.93 to 0.97. Although the Smartphone Addiction Scale has been translated and validated in French in a shorter 10-statement version [15], the original 33-items scale has not been translated and validated in French.

The Nomophobia Questionnaire was designed by Yildirim and Correia [7] and measures individuals' fear of being separated from their smartphones. It measures 4 dimensions of smartphone separation fear: not being able to communicate, losing connection with the outside world, not being able to access information, and giving up comfort. According to the research work of Jahrami et al [16], this questionnaire has been translated into 7 languages (eg, Chinese [17], Italian [18]) and consistently has very good internal consistency ($0.88 \leq \alpha \leq 0.96$).

The Problematic Use of Mobile Phones scale, designed by Merlo et al [8], draws heavily on the criteria for online gaming addiction in DSM versions IV [10] and V [2]. This scale considers 10 dimensions of problematic smartphone use in young adults: tolerance, withdrawal symptoms, smartphone use longer than expected, time spent on the smartphone, irrepressible desire to use the smartphone, abandonment or reduction of other activities, smartphone use despite physical and psychological consequences, smartphone use in risky situations, smartphone use despite social consequences, and inability to fulfill obligations due to use. Although being less validated, this scale has been translated into English [8], German [19], and Arabic [20], and its consistency appears very satisfactory ($0.90 \leq \alpha \leq 0.94$).

To date, 2 other scales measuring problematic smartphone use by young adults have been validated in French. The first scale, that is, the Implicit Association Test (smartphone) [21] is strongly inspired by the internet addiction test developed by Young [22] and not according to a set of diagnostic criteria for problematic smartphone use. However, the factor structures and internal consistency of the scale ($\alpha=0.93$) were good. The other scale, that is, the Problematic Mobile Phone Use Questionnaire by Lopez-Fernandez et al [23] is divided into 3 categories and has good internal consistency: dangerous use ($\alpha=0.81$), prohibited use ($\alpha=0.74$), and dependent use ($\alpha=0.90$). The strength of this scale lies in its cross-cultural validity (eg, French, German, Hungarian, English, Finnish, Italian, Polish, Spanish), but surprisingly, it is one of the few scales that is not based on any version of DSM. However, recent research has shown that taking inspiration from gaming disorder criteria to measure problematic smartphone use is a way to better understand this phenomenon [24].

The Smartphone Addiction Proneness Scale by Kim et al [9] is inspired by the assessment criteria of Young's [22] Internet Addiction Test scale. To our knowledge, it is one of the few scales designed to measure problematic smartphone use in adolescents. The Smartphone Addiction Proneness Scale assesses 4 main dimensions: craving symptoms, tolerance, predominance of virtual life, and disruption of coping in daily life. To date, it has been translated into Korean (ie, translated from English) [9], Malaysian [25], German [26], and Chinese [27]. The interrater internal consistency of this Korean scale is consistently acceptable ($\alpha=0.88$).

Given the importance of completing proposals for measuring problematic smartphone use based primarily on the findings of DSM and to better understand this societal issue, the aim of this study is to linguistically translate the Smartphone Addiction Scale, the Nomophobia Questionnaire, the Problematic Use of Mobile Phones scale, and the Smartphone Addiction Proneness Scale to French so that they are conceptually equivalent to their original versions. The cross-cultural adaptation process will be performed using the forward/backward method, a commonly used method [28-32] and recommended by the World Health Organization [33]. The translation will be considered valid when the tests performed by the participants are considered conclusive.

This step is a prerequisite for the quantitative validation of the questionnaires.

Methods

Background

To assess the conceptual equivalence of the questionnaires, that is, whether the items in the original language have a similar meaning to the French version, we opted for the forward/backward translation method. This is the most commonly used technique for cross-cultural research, and we followed the recommendations of Epstein et al [34]. We have decided to validate these instruments for use exclusively in France.

Ethics Approval

This study was approved by the Grenoble Alpes University Hospital Center and has received ethics approval from the South-East I Ethics Committee for the Protection of Individuals (approval 2022-A01943-40).

Questionnaires

The 4 questionnaires targeted for the cross-cultural procedure were as follows. The Smartphone Addiction Scale [6] contains 33 items, and the response scale ranges from 1 (strongly disagree) to 6 (strongly agree). Each participant's total score can range from 33 to 188 points, and a higher score indicates more problematic smartphone use. The Nomophobia Questionnaire [7] contains 20 items, and the response scale ranges from 1 (strongly disagree) to 7 (strongly agree); each participant's total score can therefore vary from 20 to 140 points, and a higher score indicates more problematic smartphone use. The Problematic Use of Mobile Phones scale contains 20 items [8]. The response scale ranges from 1 (not at all in agreement) to 5 (totally in agreement), and each participant's total score can vary from 20 to 100 points, and a higher score indicates more problematic smartphone use. Finally, the Smartphone Addiction Proneness Scale by Kim et al [9] contains 15 items. The response scale ranges from 1 (not at all in agreement) to 4 (totally in agreement), and each participant's total score can vary from 15 to 60 points, and a higher score indicates more problematic smartphone use.

Forward Translation

The first translation phase involved recruiting independent translators who were both 2 native French speakers bilingual in English and 2 native English speakers bilingual in French and asking them to translate the original English version into French. It was recommended that at least one of the translators knows the concept of the questionnaires to be measured and that at least one of them does not know the objective of the questionnaire. In this study, 4 qualified translators were involved: 2 translators were familiar with the concept of problematic smartphone use, while the other 2 were not. Two of the experts had backgrounds in health and psychology research: one was a novice and the other had prior experience in translation. At the end of the 4 translations, a single version was obtained after a reconciliation meeting of the 4 translators. Each questionnaire was checked for errors in spelling, grammar, punctuation, and translation of terminology and style against the original English version.

Backward Translation

In the second phase, the French versions were retranslated into English by a native English speaker who had no clinical or medical expertise. The instructions that were given to the native English speaker were to translate each questionnaire literally. By comparing the translated English versions with the original questionnaires, the French translations were modified to be consistent with the originals. This step ensured the accuracy of the translations and highlighted any phrasing that could cause confusion.

Linguistic Verification

In the third phase, 2 professionals familiar with the concept of problematic smartphone use were asked to comment and suggest modifications to the statements if necessary. We followed a universalist approach for equivalence, assessing conceptual, item, semantic, operational, measurement, and functional equivalence [35].

Final Verification

The final verification phase involved performing a preliminary pilot test with participants involved in the problem being measured to assess the clarity, intelligibility, and acceptability of the translations. Between December 2022 and June 2023, 18 French native speaker participants—10 students older than 18 years (ie, 6 girls and 4 boys) and 8 adolescents aged 12-17 years (ie, 6 girls and 2 boys)—completed the questionnaires for the final version during face-to-face interviews. Each interview lasted a maximum of 30 minutes. For minor participants, consent to take part in the study was obtained from the parents, while adult participants gave their own consent. Participants could be in good health or experiencing any medical condition and they could come from rural or urban areas in France. During the interviews, a researcher asked the participants questions such as "Did you understand the instructions, the items, and the response scales?"; "How did you interpret this item?"; or "What would you suggest as a reformulation?" The principal investigator checked the proofs of the final version and corrected any errors.

Data Analysis

To analyze the qualitative data from the translation process, we employed a thematic approach focused on assessing the conceptual equivalence of the translated items. Responses from the participants during the individual interviews were transcribed and systematically analyzed to identify elements that could pose comprehension issues. Particular attention was given to semantic discrepancies between the original and translated versions as well as suggestions for reformulation provided by the participants. The data were then coded according to the types of mistranslations identified (eg, item comprehension, appropriateness of terms used) and grouped into recurring themes. This approach enabled us to evaluate the clarity and acceptability of the translations as well as to adjust the problematic items.

Results

Characteristics of the Translators and the Participants

The age, gender, and specific cultural characteristics of the translators and the student and adolescent participants are detailed in Table 1.

Table 1. Age, gender, and cultural characteristics of the translators and the participants.

Translation type	Translators/participants (n)	Mean age (years)	Gender	Cultural characteristics
Forward translation	4 translators	25.5 (range 20-31)	2 females, 2 males	• 2 native English, 2 native French • 2 were familiar with the concept of problematic smartphone use (eHealth and psychology) • 2 were not familiar with the concept (economics and education) • 2 were from the south of France and 2 were from the north
Backward translation	1 translator	28	1 female	• Native English • No clinical or medical expertise
Linguistic verification	2 experts	27.5 (range 24-31)	1 female, 1 male	• 2 native French
Final verification				
	8 adolescents	14.5 (range 12-17)	6 females, 2 males	• Good health or experiencing any medical condition • 5 from rural areas and 3 from urban areas in France
	10 students	19.8 (range 18-24)	6 females, 4 males	• Good health or experiencing any medical condition • 6 students in health sciences, 3 students in psychology, 1 student in philosophy

Forward Translation

The 4 translators translated the titles, dimensions, instructions, response modalities, and statements of each questionnaire. These 4 translations were synthesized into a consensual version A. Concerning questionnaire dimensions, a few terms were more difficult to translate from English to French. For example, the dimension “withdrawal” of the Smartphone Addiction Scale, the Problematic Use of Mobile Phones scale, and the Smartphone Addiction Proneness Scale was translated as “sevrage.” For the Nomophobia Questionnaire, 1 dimension was more challenging to translate into French: “giving up convenience.” The term “convenience” can have several meanings in French, but its translation as “confort” (comfort) appears to be the most appropriate, especially when defining this dimension as the loss of the comfort provided by smartphone use. The concept of “craving” for the Nomophobia Questionnaire is particularly complex to define in French as it does not have a direct equivalent. In clinical terms, craving corresponds to a strong impulsive desire to use one’s object of addiction, which in our context is the respondent’s smartphone use. Therefore, in French, we have defined this dimension as “désir” (desire). For the other dimensions, the terms used in French were mostly similar.

For the items in the Smartphone Addiction Scale, rather than using the infinitive form of verbs as proposed in the original version, we translated them by systematically beginning the verb with the first-person singular conjugation. For example, for the original item “missing planned work due to smartphone use,” we translated it as “J’oublie du travail planifié à cause de l’utilisation du smartphone.” This will help participants better understand the statements. Three statements were more difficult to translate into French: “feeling pleasant or excited while using a smartphone” and “feeling impatient and fretful when I am not holding my smartphone.” The term “pleasant” was suggested in French as “calme” (calm), and since the term “impatient” does not have a direct equivalent in French, we propose the homograph “impatient.” The verb “to annoy” present in 2 items of the Nomophobia Questionnaire required more debate regarding its translation. This term can imply constraint, boredom, or irritation depending on the context; therefore, we chose to translate it as “contrariété” (contrariety) to preserve the statement and its dimension without distorting it. The statement “using a smartphone is more enjoyable than spending time with family or friends” from the Smartphone Addiction Proneness Scale also sparked discussions among translators because not everyone agreed on how to translate “enjoyable.” This term can signify a pleasant or amusing moment depending on the context. Here, we opted for the most common translation, “agréable” (pleasant). For the other items, the differences in translation were minor and did not alter the literal meaning of the statements.

Backward Translation

Following this first translation, a native English translator backtranslated the questionnaires from the consensus version A into an English version B. This version B was compared to the original questionnaire in English. For the Smartphone

Addiction Scale, 18 items were perfectly backtranslated, while 14 were slightly modified, but this did not alter their meaning and understanding. One item was retranslated, changing its meaning: “missing planned work due to smartphone use” was backtranslated as “I use my smartphone longer than planned.” We reworked this item in French to maintain its original meaning as “Je manque du travail planifié à cause de l’utilisation du smartphone.” For the Problematic Use of Mobile Phones scale, 17 statements were perfectly backtranslated, and the others underwent slight modifications without impacting their meaning and understanding. Backtranslation was even better for the Smartphone Addiction Proneness Scale, with 13 out of 15 items perfectly backtranslated and 2 having slight modifications without consequence. The backtranslation for the Nomophobia Questionnaire was satisfactory, with 10 out of 20 items varying slightly in wordings but without altering their meaning. The others were perfectly backtranslated. Thus, we obtained a translated version C for the 4 questionnaires.

Linguistic Verification

The 2 researchers responsible for linguistic verification submitted 4 suggestions for conceptual modifications for the Smartphone Addiction Scale, 3 for the Nomophobia Questionnaire, and 1 for the Problematic Use of Mobile Phones scale, some of which were incorporated into the final C version. No additional modifications were suggested for the Smartphone Addiction Proneness Scale.

Final Verification

In this final step, feedback from adolescents and students was minimal; therefore, we produced a final version D based on their input. Overall, they understood all the instructions, items, and response modalities well. Participants encountered no comprehension problems except for 3 items. For the Nomophobia Questionnaire, 2 students reported a comprehension difficulty with item 18. In English, it refers to checking “online connections and networks,” and we translated it as “online notifications and networks.” However, “online networks” made less sense than “social networks” for these students. We therefore modified the item. In addition, item 6 of the Problematic Use of Mobile Phones scale was difficult for one student to understand, as the translators had translated it by mixing the present and past tenses. We therefore modified the item from “J’ai pensé par le passé qu’il n’était pas normal de passer autant de temps sur le smartphone comme je le fais” to “Il m’est arrivé de penser qu’il n’était pas normal de passer autant de temps que moi à utiliser un smartphone.” In this way, students gained a better understanding of the item. Finally, item 22 of the Smartphone Addiction Scale had to be clarified, as 2 participants did not understand what “losing a friend” meant. In fact, the meaning of loss is important in French, as it could mean literally losing a friend through death or signify a friendship breakup. We have therefore clarified at the end of the item that this refers to friendship breakup in this context. The final translated versions are listed in Tables 2-5.

Table 2. Problematic Use of Mobile Phones scale translation.

Item number	Item description	Response methods				
		1[a]	2	3	4	5[b]
1	ET[c]: When I decrease the amount of time spent using my cell phone I feel less satisfied. FT[d]: Quand je réduis le temps passé sur mon smartphone, je me sens moins satisfait·e.					
2	ET: I need more time using my cell phone to feel satisfied than I used to need. FT: J'ai besoin de plus de temps d'utilisation de mon smartphone afin de me sentir aussi satisfait·e qu'auparavant.					
3	ET: When I stop using my cell phone, I get moody and irritable. FT: Quand j'arrête d'utiliser mon smartphone, je deviens de mauvaise humeur et irritable.					
4	ET: It would be very difficult emotionally to give up my cell phone. FT: Ce serait difficile émotionnellement de renoncer à mon smartphone.					
5	ET: The amount of time I spend using my cell phone keeps me from doing other important work. FT: Le temps passé sur mon smartphone m'empêche de faire d'autres tâches importantes.					
6	ET: I have thought in the past that it is not normal to spend as much time using a cell phone as I do. FT: Il m'est arrivé·e de penser qu'il n'était pas normal de passer autant de temps que moi à utiliser un smartphone.					
7	ET: I think I might be spending too much time using my cell phone. FT: Je pense que je passe trop de temps sur mon smartphone.					
8	ET: People tell me I spend too much time using my cell phone. FT: Les gens me disent que je passe trop de temps sur mon smartphone.					
9	ET: When I am not using my cell phone, I am thinking about using it or planning the next time I can use it. FT: Quand je n'utilise pas mon smartphone, je pense à l'utiliser ou à prévoir ma prochaine utilisation.					
10	ET: I feel anxious if I have not received a call or message in some time. FT: Je me sens anxieux·se si je n'ai pas reçu d'appel ou de message depuis un certain moment.					
11	ET: I have ignored the people I'm with in order to use my cell phone. FT: J'ai ignoré les personnes avec qui j'étais pour utiliser mon smartphone.					
12	ET: I have used my cell phone when I knew I should be doing work/schoolwork. FT: J'ai utilisé mon téléphone alors que je savais que je devais travailler/faire des devoirs.					
13	ET: I have used my cell phone when I knew I should be sleeping. FT: J'ai déjà utilisé mon smartphone alors que je savais que je devais dormir.					
14	ET: When I stop using my cell phone because it is interfering with my life, I usually return to it. FT: Quand j'arrête d'utiliser mon smartphone parce qu'il interfère avec ma vie, je finis généralement par l'utiliser à nouveau.					
15	ET: I have gotten into trouble at work or school because of my cell phone use. FT: J'ai eu des problèmes au travail ou à l'école à cause de l'utilisation de mon smartphone.					
16	ET: At times, I find myself using my cell phone instead of spending time with people who are important to me and want to spend time with me. FT: Parfois, je me retrouve à utiliser mon smartphone au lieu de passer du temps avec des personnes qui sont importantes pour moi et qui souhaitent passer du temps avec moi.					
17	ET: I have used my cell phone when I knew it was dangerous to do so. FT: J'ai utilisé mon téléphone alors que je savais qu'il était dangereux de le faire.					
18	ET: I have almost caused an accident because of my cell phone use. FT: J'ai presque causé un accident à cause de l'utilisation de mon smartphone.					
19	ET: My cell phone use has caused me problems in a relationship. FT: L'utilisation de mon smartphone a causé des problèmes dans une relation.					
20	ET: I have continued to use my cell phone even when someone asked me to stop. FT: J'ai continué d'utiliser mon smartphone, même quand quelqu'un m'a demandé d'arrêter.					

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: strongly agree; French translation: fortement en accord.

[c]ET: English translation.

[d]FT: French translation.

Table 3. Smartphone Addiction Scale translation.

Item number	Item description	Response methods					
		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
1	ET[g]: Missing planned work due to smartphone use. FT[h]: Je manque du travail planifié à cause de l'utilisation du smartphone.						
2	ET: Having a hard time concentrating in class while doing assignments or while working due to smartphone use. FT: J'ai des difficultés à me concentrer en classe, pendant mes devoirs ou pendant le travail à cause de l'utilisation du smartphone.						
3	ET: Experiencing lightheadedness or blurred vision due to excessive smartphone use. FT: Je ressens des vertiges ou une vision floue à cause de l'utilisation excessive du smartphone.						
4	ET: Feeling pain in the wrist or at the back of the neck while using a smartphone. FT: Je ressens de la douleur dans les poignets ou derrière le cou pendant que j'utilise un smartphone.						
5	ET: Feeling tired and lacking adequate sleep due to excessive smartphone use. FT: Je me sens fatigué et en manque de sommeil suffisant à cause de l'utilisation excessive du smartphone.						
6	ET: Feeling calm or cozy while using a smartphone. FT: Je me sens calme et réconforté·e quand j'utilise un smartphone.						
7	ET: Feeling pleasant or excited while using a smartphone. FT: Je me sens calme ou excité·e en utilisant le smartphone.						
8	ET: Feeling confident while using a smartphone. FT: Je me sens confiant·e en utilisant le smartphone.						
9	ET: Being able to get rid of stress with a smartphone. FT: Je suis capable de me débarrasser du stress avec un smartphone.						
10	ET: There is nothing more fun to do than using my smartphone. FT: Il n'y a rien de plus amusant que d'utiliser mon smartphone.						
11	ET: My life would be empty without my smartphone. FT: Ma vie serait vide sans mon smartphone.						
12	ET: Feeling most liberal while using a smartphone. FT: Je me sens libre quand j'utilise un smartphone.						
13	ET: Using a smartphone is the most fun thing to do. FT: Utiliser un smartphone est la chose la plus amusante à faire.						
14	ET: Won't be able to stand not having a smartphone. FT: Je ne supporterais pas de ne pas avoir de smartphone.						
15	ET: Feeling impatient and fretful when I am not holding my smartphone. FT: Je me sens impatient·e et irrité·e quand je ne tiens pas mon smartphone.						
16	ET: Having my smartphone in my mind even when I am not using it. FT: Je pense à mon smartphone même lorsque je ne l'utilise pas.						
17	ET: I will never give up using my smartphone even when my daily life is already greatly affected by it. FT: Je ne renoncerais jamais à utiliser mon smartphone, même si ma vie quotidienne en est déjà fortement affectée.						
18	ET: Getting irritated when bothered while using my smartphone. FT: Je me sens irrité·e quand on me dérange pendant que j'utilise mon smartphone.						
19	ET: Bringing my smartphone to the toilet even when I am in a hurry to get there. FT: J'apporte mon smartphone aux toilettes même lorsque je suis pressé·e d'y aller.						
20	ET: Feeling great meeting more people via smartphone use. FT: Je me sens bien en rencontrant plus de personnes par le biais de l'utilisation de mon smartphone.						

Item number	Item description	Response methods					
		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
21	ET: Feeling that my relationships with my smartphone buddies are more intimate than my relationships with my real-life friends. FT: J'ai l'impression que mes relations virtuelles sont plus intimes que mes relations avec mes amis de la vie réelle.						
22	ET: Not being able to use my smartphone would be as painful as losing a friend. FT: Ne pas pouvoir utiliser mon smartphone serait aussi douloureux que de perdre un ami (rupture amicale).						
23	ET: Feeling that my smartphone buddies understand me better than my real-life friends. FT: J'ai l'impression que mes amis virtuels me comprennent mieux que mes amis de la vie réelle.						
24	ET: Constantly checking my smartphone so as not to miss conversations between other people on Twitter or Facebook. FT: Je vérifie constamment mon smartphone afin de ne manquer aucune conversation entre d'autres personnes sur Twitter ou Facebook.						
25	ET: Checking social networking service sites like Twitter or Facebook right after waking up. FT: Je vérifie les réseaux sociaux comme Twitter ou Facebook juste après le réveil.						
26	ET: Preferring talking with my smartphone buddies to hanging out with my real-life friends or with the other members of my family. FT: Je préfère discuter avec mes amis virtuels plutôt que passer du temps avec mes amis de la vie réelle ou avec les membres de ma famille.						
27	ET: Preferring searching from my smartphone to asking other people. FT: Je préfère faire mes recherches à partir de mon smartphone plutôt que de demander à d'autres personnes.						
28	ET: My fully charged battery does not last for one whole day. FT: Ma batterie pleinement chargée ne dure pas une journée entière.						
29	ET: Using my smartphone longer than I had intended. FT: J'utilise mon smartphone plus longtemps que prévu.						
30	ET: Feeling the urge to use my smartphone again right after I stopped using it. FT: Je ressens le besoin d'utiliser à nouveau mon smartphone juste après avoir arrêté de l'utiliser.						
31	ET: Having tried time and again to shorten my smartphone use time but failing all the time. FT: J'ai essayé à plusieurs reprises de réduire le temps d'utilisation de mon smartphone mais j'échoue à chaque fois.						
32	ET: Always thinking that I should shorten my smartphone use time. FT: Je pense toujours que je devrais réduire mon utilisation du smartphone.						
33	ET: The people around me tell me that I use my smartphone too much. FT: Les gens autour de moi me disent que j'utilise trop mon smartphone.						

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: disagree; French translation: en désaccord.

[c]English translation: weakly disagree; French translation: un peu en désaccord.

[d]English translation: weakly agree; French translation: un peu en accord.

[e]English translation: agree; French translation: en accord.

[f]English translation: strongly agree; French translation: fortement en accord.

[g]ET: English translation.

[h]FT: French translation.

Table 4. Nomophobia Questionnaire translation.

Item number	Item description	Response methods						
		1[a]	2	3	4	5	6	7[b]
1	ET[c]: I would feel uncomfortable without constant access to information through my smartphone. FT[d]: Je me sentirais mal à l'aise sans l'accès constant à l'information au travers de mon smartphone.							
2	ET: I would be annoyed if I could not look information up on my smartphone when I wanted to do so. FT: Je me sentirais contrarié·e si je ne pouvais pas chercher de l'information sur mon smartphone quand je le souhaite.							
3	ET: Being unable to get the news (eg, happenings, weather, etc) on my smartphone would make me nervous. FT: Être incapable d'avoir accès aux nouvelles (eg, actualités, météo, etc.) sur mon smartphone me rendrait nerveux·se.							
4	ET: I would be annoyed if I could not use my smartphone and/or its capabilities when I wanted to do so. FT: Je serais contrarié·e si je ne pouvais pas utiliser mon smartphone et/ou ses fonctionnalités quand je le souhaite.							
5	ET: Running out of battery in my smartphone would scare me. FT: Ne plus avoir de batterie sur mon smartphone me ferait peur.							
6	ET: If I were to run out of credits or hit my monthly data limit, I would panic. FT: Si je me retrouvais sans crédits ou si j'atteignais ma limite de données mensuelle, je paniquerais.							
7	ET: If I did not have a data signal or could not connect to Wi-Fi, then I would constantly check to see if I had a signal or could find a Wi-Fi network. FT: Si je n'avais pas de signal réseau ou ne pouvais pas me connecter au Wi-Fi, je vérifierais constamment si j'ai du réseau ou si je peux trouver un réseau Wi-Fi.							
8	ET: If I could not use my smartphone, I would be afraid of getting stranded somewhere. FT: Si je ne pouvais pas utiliser mon smartphone, j'aurais peur de rester bloqué·e quelque part.							
9	ET: If I could not check my smartphone for a while, I would feel a desire to check it. FT: Si je ne pouvais pas consulter mon smartphone pendant un certain temps, j'aurais envie de le consulter.							
ET: If I did not have my smartphone with me, FT: Si je n'avais pas mon smartphone avec moi,								
10	ET: I would feel anxious because I could not instantly communicate with my family and/or friends. FT: Je me sentirais anxieux·se parce que je ne pourrais pas communiquer instantanément avec ma famille et/ou mes amis.							
11	ET: I would be worried because my family and/or friends could not reach me. FT: Je serais inquiet·e parce que ma famille et/ou mes amis ne pourraient pas me joindre.							
12	ET: I would feel nervous because I would not be able to receive text messages and calls. FT: Je me sentirais nerveux·se de ne pas pouvoir recevoir des messages ou des appels.							
13	ET: I would be anxious because I could not keep in touch with my family and/or friends. FT: Je me sentirais anxieux·se parce que je ne pourrais pas rester en contact avec ma famille et/ou mes amis.							
14	ET: I would be nervous because I could not know if someone had tried to get a hold of me. FT: Je serais nerveux·se parce que je ne pourrais pas savoir si quelqu'un a essayé de me contacter.							
15	ET: I would feel anxious because my constant connection to my family and friends would be broken. FT: Je me sentirais anxieux·se parce que ma connexion constante avec ma famille et mes amis serait rompue.							
16	ET: I would be nervous because I would be disconnected from my online identity. FT: Je serais nerveux·se parce que je serais déconnecté·e de mon identité en ligne.							
17	ET: I would be uncomfortable because I could not stay up-to-date with social media and online networks. FT: Je serais mal à l'aise parce que je ne pourrais pas être à jour avec les réseaux sociaux ou les réseaux en ligne.							

Item number	Item description	Response methods						
		1[a]	2	3	4	5	6	7[b]
18	ET: I would feel awkward because I could not check my notifications for updates from my connections and online networks. FT: Je me sentirais mal à l'aise parce que je ne pourrais pas vérifier mes notifications et réseaux sociaux.							
19	ET: I would feel anxious because I could not check my email messages. FT: Je me sentirais anxieux·se parce que je ne pourrais pas vérifier mes emails.							
20	ET: I would feel weird because I would not know what to do. FT: Je me sentirais mal à l'aise car je ne saurais pas quoi faire.							

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: strongly agree; French translation: fortement en accord.

[c]ET: English translation.

[d]FT: French translation.

Table 5. Smartphone Addiction Proneness Scale translation.

Item number	Item description	Response methods			
		1[a]	2[b]	3[c]	4[d]
1	ET[e]: My school grades dropped due to excessive smartphone use. FT[f]: Mes notes scolaires ont baissé à cause d'une utilisation excessive du smartphone.				
2	ET: I have a hard time doing what I have planned (study, do homework, or go to afterschool classes) due to using smartphone. FT: J'ai du mal à faire ce que j'avais prévu (étudier, faire mes devoirs, ou aller aux activités scolaires) à cause de l'utilisation du smartphone.				
3	ET: People frequently comment on my excessive smartphone use. FT: Les gens commentent fréquemment mon utilisation excessive du smartphone.				
4	ET: Family or friends complain that I use my smartphone too much. FT: Ma famille ou mes amis se plaignent que j'utilise trop mon smartphone.				
5	ET: My smartphone does not distract me from my studies (reversed item). FT: Mon smartphone ne me distrait pas de mes études. (énoncé inversé)				
6	ET: Using a smartphone is more enjoyable than spending time with family or friends. FT: Utiliser un smartphone est plus agréable que de passer du temps avec ma famille ou mes amis.				
7	ET: When I cannot use a smartphone, I feel like I have lost the entire world. FT: Quand je ne peux pas utiliser un smartphone, j'ai l'impression d'avoir perdu le monde entier.				
8	ET: It would be painful if I am not allowed to use a smartphone. FT: Ce serait douloureux si je n'étais pas autorisé·e à utiliser un smartphone.				
9	ET: I get restless and nervous when I am without a smartphone. FT: Je deviens agité·e et nerveux·se quand je suis sans smartphone.				
10	ET: I am not anxious even when I am without a smartphone (reversed item). FT: Je ne suis pas anxieux·se, même quand je suis sans smartphone. (énoncé inversé)				
11	ET: I panic when I cannot use my smartphone. FT: Je panique quand je ne peux pas utiliser mon smartphone.				
12	ET: I try cutting my smartphone usage time, but I fail. FT: J'essaie de réduire le temps d'utilisation de mon smartphone, mais j'échoue.				
13	ET: I can control my smartphone usage time (reversed item). FT: Je peux contrôler le temps d'utilisation de mon smartphone. (énoncé inversé)				
14	ET: Even when I think I should stop, I continue to use my smartphone too much. FT: Même lorsque je pense que je devrais arrêter, je continue de trop utiliser mon smartphone.				
15	ET: Spending a lot of time on my smartphone has become a habit. FT: Passer beaucoup de temps sur mon smartphone est devenu une habitude.				

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: disagree; French translation: en désaccord.

[c]English translation: agree; French translation: en accord.

[d]English translation: strongly agree; French translation: fortement en accord.

[e]ET: English translation.

[f]FT: French translation.

Discussion

Principal Findings

The Problematic Use of Mobile Phones scale, the Smartphone Addiction Scale, the Smartphone Addiction Proneness Scale, and the Nomophobia Questionnaire for young adult students and adolescents have been developed and validated in English to measure problematic smartphone use, but they are not available in French. In this study, we provide their translations and linguistic validation in French.

Translating questionnaires into other languages can be complex, as terms in one language can be interpreted in different ways in another, leading to errors in meaning. For this reason, we

have followed the rigorous and standardized steps of linguistic validation defined by the forward/backward method [34].

The results of the translations showed that after 4 stages of translation, the content of the French questionnaires was identical to that of the English ones. The translators, adolescents, and young adults from different educational/professional and geographical backgrounds did not reveal any differences in understanding, suggesting that the cultural adaptation of the questionnaires was respected. These translations suggest that cross-cultural adaptation can be used for comparative purposes in other languages such as English, Portuguese, Arabic, or German.

However, 3 limitations need to be addressed in this study. First, the number of adolescents and young adult participants was less than 20. Although it would be more robust to interview more participants, the minimum number of participants for linguistic validation is not clearly defined [35]. Moreover, the adolescents and young adults provided only few comments on the clarity and understanding of the questionnaires. Second, although 4 volunteer translators were consulted, a more diversified panel of experts would have enabled us to avoid any cultural bias in the translation of the questionnaires. We were able to bring together 4 translators of different genders (ie, 2 men and 2 women) and from different disciplines (ie, eHealth, psychology, education, and economics) who did not have professional knowledge of the concept of problematic smartphone use. Despite these considerations, France is a country with a rich cultural diversity, which may include some regional and linguistic differences. We could have included native experts from different regions with marked linguistic peculiarities (eg, Brittany, Corsica). It is important to note that the 4 translators selected came from different regions of France, which ensures that we have reduced at least some of the cultural biases in translation. The last limitation was that this study focuses solely on linguistic validation without demonstrating quantitative validation. According to Terwee and colleagues [36], the translation process is the first necessary step for validation, but it is essential to verify content validity, internal consistency, criterion validity, construct validity, and reproducibility. This is particularly important here because problematic smartphone use, as measured by the questionnaires, is often assessed across multiple dimensions such as tolerance, withdrawal symptoms, overuse, or physical and psychological consequences [3]. These validation steps, especially regarding psychometric properties, are essential and must be completed before any use of these questionnaires in French. They can then be reliably used to better understand problematic smartphone use among French speakers and for conducting cross-cultural comparisons. These questionnaires will be relevant for gaining a better understanding of how problematic smartphone use is characterized among French adults and adolescents and can be utilized for screening and potentially monitoring patients with conditions associated with problematic smartphone use [5,37].

Conclusion

We succeeded in adapting and effectively translating 4 questionnaires assessing problematic smartphone use in a French-speaking context. This step remains a prerequisite to validating the questionnaires, which can serve as valuable research instruments for investigating and addressing issues related to problematic smartphone use in French-speaking countries and for making international comparisons. Given that problematic smartphone use is an emerging concern with significant health impacts [1], it is imperative that researchers and health care professionals use reliable measurement instruments.

Acknowledgments

This study would not have been possible without the support of Université Grenoble Alpes, Grenoble Alpes University Hospital Center, and LINKSIUM (Society of Technology Transfer and Startup, Grenoble Alpes). This work was partially funded by MIAI Cluster (ANR-23-IACL-0006) and the French National Research Agency (Agence Nationale de la Recherche) within the framework of the “Investissements d’avenir” program (ANR-10-AIRT-05, 2012; ANR-15-IDEX-02, 2016). The funders played no role in the study design, data collection, analysis and interpretation of data, or writing of this manuscript.

Authors' Contributions

IEB collected data, reviewed and assessed the results, wrote the original draft, wrote and reviewed the revised version of the manuscript, and acquired funding. MR collected analyzed and interpreted data, wrote the original draft, wrote and reviewed the revised version of the manuscript, and acquired funding. AB conceptualized and designed the study, reviewed and assessed the results, wrote and reviewed the revised version of the manuscript. NV conceptualized and designed the study, supervised the project, reviewed and assessed the results, wrote and reviewed the revised version of the manuscript and acquired funding. All authors reviewed and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Wacks Y, Weinstein AM. Excessive smartphone use is associated with health problems in adolescents and young adults. Front Psychiatry 2021;12:669042 [FREE Full text] [doi: 10.3389/fpsyt.2021.669042] [Medline: 34140904]

2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington DC: American Psychiatric Association Publishing; May 18, 2013.
3. Harris B, Regan T, Schueler J, Fields SA. Problematic mobile phone and smartphone use scales: a systematic review. Front Psychol 2020;11:672 [FREE Full text] [doi: 10.3389/fpsyg.2020.00672] [Medline: 32431636]
4. Notara V, Vagka E, Gnardellis C, Lagiou A. The emerging phenomenon of nomophobia in young adults: a systematic review study. Addict Health 2021 Apr;13(2):120-136 [FREE Full text] [doi: 10.22122/ahj.v13i2.309] [Medline: 34703533]
5. Ratan ZA, Parrish A, Zaman SB, Alotaibi MS, Hosseinzadeh H. Smartphone addiction and associated health outcomes in adult populations: a systematic review. Int J Environ Res Public Health 2021 Nov 22;18(22):12257 [FREE Full text] [doi: 10.3390/ijerph182212257] [Medline: 34832011]
6. Kwon M, Lee J, Won W, Park J, Min J, Hahn C, et al. Development and validation of a smartphone addiction scale (SAS). PLoS One 2013;8(2):e56936 [FREE Full text] [doi: 10.1371/journal.pone.0056936] [Medline: 23468893]
7. Yildirim C, Correia A. Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. Computers in Human Behavior 2015 Aug;49:130-137. [doi: 10.1016/j.chb.2015.02.059]
8. Merlo LJ, Stone AM, Bibbey A. Measuring problematic mobile phone use: development and preliminary psychometric properties of the PUMP scale. J Addict 2013;2013:912807 [FREE Full text] [doi: 10.1155/2013/912807] [Medline: 24826371]
9. Kim D, Lee Y, Lee J, Nam JK, Chung Y. Development of Korean smartphone addiction proneness scale for youth. PLoS One 2014;9(5):e97920 [FREE Full text] [doi: 10.1371/journal.pone.0097920] [Medline: 24848006]
10. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington DC: American Psychiatric Publishing; May 15, 1994.
11. Sfendla A, Laita M, Nejjar B, Souirti Z, Touhami AAO, Senhaji M. Reliability of the Arabic smartphone addiction scale and smartphone addiction scale-short version in two different Moroccan samples. Cyberpsychol Behav Soc Netw 2018 May;21(5):325-332. [doi: 10.1089/cyber.2017.0411] [Medline: 29762065]
12. Andrade ALM, Kim D, Scatena A, Enes CC, Enumo SRF, De Micheli D. Validity and reliability of the Brazilian version of the smartphone addiction scale-long version (SAS-LV). Trends in Psychol 2020 Dec 01;29(2):302-319. [doi: 10.1007/s43076-020-00046-y]
13. Ching SM, Yee A, Ramachandran V, Sazlly Lim SM, Wan Sulaiman WA, Foo YL, et al. Validation of a Malay version of the smartphone addiction scale among medical students in Malaysia. PLoS One 2015;10(10):e0139337 [FREE Full text] [doi: 10.1371/journal.pone.0139337] [Medline: 26431511]
14. Mokhtarinia H, Rafinia M, Shahbazi A, Khodaie Ardakani MR, Ghaedamini Harouni GR, Philip Gabel C. The cross-cultural adaptation of the smartphone addiction scale to Persian. Iran Rehabil J 2020 Mar 30:91-98. [doi: 10.32598/irj.18.1.919.1]
15. Lopez-Fernandez O. Short version of the smartphone addiction scale adapted to Spanish and French: Towards a cross-cultural research in problematic mobile phone use. Addict Behav 2017 Jan;64:275-280. [doi: 10.1016/j.addbeh.2015.11.013] [Medline: 26685805]
16. Jahrami H, Saif Z, Trabelsi K, Bragazzi NL, Vitiello MV. Internal consistency and structural validity of the nomophobia questionnaire (NMP-Q) and its translations: A systematic review with meta-analysis. Heliyon 2023 Apr;9(4):e15464 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e15464] [Medline: 37128316]
17. Gao Y, Dai H, Jia G, Liang C, Tong T, Zhang Z, et al. Translation of the Chinese version of the nomophobia questionnaire and its validation among college students: factor analysis. JMIR Mhealth Uhealth 2020 Mar 13;8(3):e13561 [FREE Full text] [doi: 10.2196/13561] [Medline: 32167480]
18. Adawi M, Bragazzi NL, Argumosa-Villar L, Boada-Grau J, Vigil-Colet A, Yildirim C, et al. Translation and validation of the nomophobia questionnaire in the Italian language: exploratory factor analysis. JMIR Mhealth Uhealth 2018 Jan 22;6(1):e24 [FREE Full text] [doi: 10.2196/mhealth.9186] [Medline: 29358161]
19. Graben K, Doering BK, Jeromin F, Barke A. Problematic mobile phone use: Validity and reliability of the Problematic Use of Mobile Phone (PUMP) Scale in a German sample. Addict Behav Rep 2020 Dec;12:100297 [FREE Full text] [doi: 10.1016/j.abrep.2020.100297] [Medline: 33364306]
20. Alshahwan H, Alosaimi F, Alyahya H, Mahyijari N, Shaik S. Arabic validation of problematic use of mobile phone scale among university students in Saudi Arabia. J Nat Sci Med 2020. [doi: 10.4103/jnsm.jnsm_9_19]
21. Barrault S, Durousseau F, Ballon N, Réveillère C, Brunault P. [Smartphone addiction: French validation of the internet addiction test-smartphone version (IAT-smartphone) and associated psychopathological features]. Encephale 2019 Feb;45(1):53-59. [doi: 10.1016/j.encep.2017.12.002] [Medline: 29397925]
22. Young KS. Internet addiction: the emergence of a new clinical disorder. CyberPsychology & Behavior 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]
23. Lopez-Fernandez O, Kuss DJ, Pontes HM, Griffiths MD, Dawes C, Justice LV, et al. Measurement invariance of the short version of the problematic mobile phone use questionnaire (PMPUQ-SV) across eight languages. Int J Environ Res Public Health 2018 Jun 08;15(6):1213 [FREE Full text] [doi: 10.3390/ijerph15061213] [Medline: 29890709]
24. Ting C, Chen Y. Smartphone addiction. In: Adolescent Addiction. England: Academic Press; 2020:215-240.
25. Chan SJ, Yeo KJ, Handayani L. Psychometric properties of the smartphone addiction proneness scale in a sample of Malaysian adolescents. IJPHS 2022 Sep 01;11(3):785. [doi: 10.11591/ijphs.v11i3.21971]

26. Randler C, Wolfgang L, Matt K, Demirhan E, Horzum MB, Beşoluk S. Smartphone addiction proneness in relation to sleep and morningness-eveningness in German adolescents. J Behav Addict 2016 Sep;5(3):465-473 [FREE Full text] [doi: 10.1556/2006.5.2016.056] [Medline: 27499228]
27. Zhang Q, Liu Z. Smartphone addiction proneness scale for adolescents? Chinese version: revision and evaluation on reliability and validity. Chinese Journal of Public Health 2021;37(7):1149-1152. [doi: 10.11847/zgggws1129783]
28. Andersson S, Granat L, Brännström M, Sandgren A. Translation, cultural adaptation, and content validation of the palliative care self-efficacy scale for use in the Swedish context. Int J Environ Res Public Health 2022 Jan 20;19(3):1143 [FREE Full text] [doi: 10.3390/ijerph19031143] [Medline: 35162163]
29. Yoon YE, Cho SY. Translation and linguistic validation of the Korean version of the Wisconsin Stone quality of life questionnaire. Int Neurourol J 2020 Mar;24(1):77-83 [FREE Full text] [doi: 10.5213/inj.1836238.119] [Medline: 32252189]
30. Poder TG, Carrier N, Mead H, Stevens KJ. Canadian French translation and linguistic validation of the child health utility 9D (CHU9D). Health Qual Life Outcomes 2018 Aug 29;16(1):168 [FREE Full text] [doi: 10.1186/s12955-018-0998-4] [Medline: 30157857]
31. Schettini A, Rillon P, Pirson G, De Coninck C. Completion of the French translation and linguistic validation of the BODY-Q. Psychiatr Danub 2020 Sep;32(Suppl 1):150-152 [FREE Full text] [Medline: 32890380]
32. Soulard J, Vaillant J, Baillet A, Vuillerme N. Translation and French linguistic validation of the bath ankylosing spondylitis functional index and the bath ankylosing spondylitis global score in patients with axial spondyloarthritis. Curr Rheumatol Rev 2023;19(4):449-454. [doi: 10.2174/1573397118666220829124234] [Medline: 36043728]
33. Translation and linguistic evaluation protocol and supporting material. WHODAS 2.0 Translation Package. URL: https://tinyurl.com/2etx7aea [accessed 2023-12-01]
34. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol 2015 Apr;68(4):435-441. [doi: 10.1016/j.jclinepi.2014.11.021] [Medline: 25698408]
35. Harachi TW, Choi Y, Abbott RD, Catalano RF, Bliesner SL. Examining equivalence of concepts and measures in diverse samples. Prev Sci 2006 Dec;7(4):359-368 [FREE Full text] [doi: 10.1007/s11121-006-0039-0] [Medline: 16845592]
36. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
37. Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: A systematic review and meta-analysis. Psychiatry Res 2020 Feb;284:112686. [doi: 10.1016/j.psychres.2019.112686] [Medline: 31757638]

Abbreviations

DSM: Diagnostic and Statistical Manual of Mental Disorders

Edited by T de Azevedo Cardoso; submitted 25.10.23; peer-reviewed by B Imbert, H Li; comments to author 09.03.24; revised version received 26.06.24; accepted 23.09.24; published 26.02.25.

Please cite as:
El Boudi I, Riant M, Bellier A, Vuillerme N
French Versions of 4 English Questionnaires on Problematic Smartphone Use: Cross-Cultural Linguistic Translation and Adaptation Study
Interact J Med Res 2025;14:e53958
URL: https://www.i-jmr.org/2025/1/e53958
doi:10.2196/53958
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org