

Original Paper

# Predicting Depressive Symptoms Using GPS-Based Regional Data in Germany With the CORONA HEALTH App During the COVID-19 Pandemic: Cross-Sectional Study

Johanna-Sophie Edler<sup>1\*</sup>, MPsych; Michael Winter<sup>2\*</sup>, PhD; Holger Steinmetz<sup>3</sup>, PhD; Caroline Cohrdes<sup>1</sup>, PD; Harald Baumeister<sup>4</sup>, Prof Dr; Rüdiger Pryss<sup>2</sup>, Prof Dr

<sup>1</sup>Mental Health Research Unit, Department of Epidemiology and Health Monitoring, Robert Koch Institute, Berlin, Germany

<sup>2</sup>Department of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

<sup>3</sup>Department of Business Administration, Trier University, Trier, Germany

<sup>4</sup>Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

\*these authors contributed equally

**Corresponding Author:**

Johanna-Sophie Edler, MPsych

Mental Health Research Unit

Department of Epidemiology and Health Monitoring

Robert Koch Institute

PO Box 650261

Berlin, 12101

Germany

Phone: 49 1723842979

Email: [johannasophie.edler@gmail.com](mailto:johannasophie.edler@gmail.com)

## Abstract

**Background:** Numerous studies have been conducted to predict depressive symptoms using passive smartphone data, mostly integrating the GPS signal as a measure of mobility. Environmental factors have been identified as correlated with depressive symptoms in specialized studies both before and during the pandemic.

**Objective:** This study combined a data-based approach using passive smartphone data to predict self-reported depressive symptoms with a wide range of GPS-based environmental factors as predictors.

**Methods:** The CORONA HEALTH app was developed for the purpose of data collection, and this app enabled the collection of both survey and passive data via smartphone. After obtaining informed consent, we gathered GPS signals at the time of study participation and evaluated depressive symptoms in 249 Android users with the Patient Health Questionnaire-9. The only GPS-based data collected were the participants' location at the time of the questionnaire, which was used to assign participants to the nearest district for linking regional sociodemographic data. Data collection took place from July 2020 to February 2021, coinciding with the COVID-19 pandemic. Using GPS data, each dataset was linked to a wide variety of data on regional sociodemographic, geographic, and economic characteristics describing the respondent's environment, which were derived from a publicly accessible database from official German statistical offices. Moreover, pandemic-specific predictors such as the current pandemic phase or the number of new regional infections were matched via GPS. For the prediction of individual depressive symptoms, we compared 3 models (ie, ridge, lasso, and elastic net regression) and evaluated the models using 10-fold cross-validation.

**Results:** The final elastic net regression model showed the highest explained variance ( $R^2=0.06$ ) and reduced the dataset from 121 to 9 variables, the 3 main predictors being current COVID-19 infections in the respective district, the number of places in nursing homes, and the proportion of fathers receiving parental benefits. The number of places in nursing homes refers to the availability of care facilities for the elderly, which may indicate regional population characteristics that influence mental health. The proportion of fathers receiving parental benefits reflects family structure and work-life balance, which could impact stress and mental well-being during the pandemic.

**Conclusions:** Passive data describing the environment contributed to the prediction of individual depressive symptoms and revealed regional risk and protective factors that may be of interest without their inclusion in routine assessments being costly.

**KEYWORDS**

depression; COVID-19; mobile phone; geographic information systems; GPS-based data; mobile applications; mental health

## Introduction

### Background

Depression is a serious and widespread disease accompanied by a high individual burden [1]. Early detection of depressive symptoms facilitates care through early intervention and can thus prevent the development of severe symptoms such as suicidal ideation [2]. Therefore, knowledge of the current incidence and rapid recognition of trends in the prevalence of depressive symptoms is of great social relevance [3], for instance, for planning treatment strategies and identifying capacities.

To identify such trends in a timely manner, population-wide (broad) data collection in real time is pivotal. Collecting survey data requires the commitment of the participants, and they have to take the time and effort to answer questions. With the proliferation of smartphones, an alternative data source is available that does not require the effort of participants. Smartphone data have been used in research on mental health [4,5] and explicitly for depressive symptoms [6,7]. Analyses based on these smartphone data often aim to predict, for instance, depressive symptoms by including all available data [4,6,8,9]. A distinction is made between active and passive data collected via smartphones. Active smartphone data involve data that require the active involvement of the smartphone user (eg, survey data collected via the smartphone), and passive smartphone data include data that can be collected without the user's involvement [4]. Passive smartphone data can be divided into activity data, social media data, and sensor data [10]. A whole body of evidence investigates the predictive power of activity data (eg, text messages sent, calls made, and screen activity), which are understood as objective measures of behavior [11]. It has been shown that some of these activity data are clearly related to depressive symptoms [12-15]. Social media data include information about posts on social networks, and these aspects of social interaction are also associated with depressive symptoms [16]. Smartphone sensor data involve the data measured with the sensors integrated in smartphones, such as GPS [10]. Studies often use GPS to analyze mobility behavior and have been able to identify correlations with depressive symptoms as well [4,7,10,14,17-20]. However, by including GPS sensor data in research on depressive symptoms, it is further possible to describe the participants' environment in detail. Environmental factors include, for example, the sociodemographics of the population living in the surrounding area, economic conditions, social affairs, or living conditions (Table S1 in [Multimedia Appendix 1](#)). These environmental factors are external factors that affect people's living conditions. For example, urbanization and population density are controversially discussed as risk factors for depressive symptoms [21]. In addition, different types of noise and air pollution were predominantly identified as positively related to depressive symptoms [21]. A poor home or building environment has been

repeatedly identified as positively correlated with depressive symptoms, and the same applies to a lack of green areas [21].

This study was conducted during the COVID-19 pandemic. This was a time when people spent more time at home due to contact restrictions and lockdowns [22]. Researchers argue that environmental factors may have been particularly important for mental health during the pandemic, as more time was spent in those same environments [23-25]. This is also indicated by the results showing an association between housing conditions (eg, available space) and depressive symptoms during the COVID-19 pandemic [23,26]. For this reason, environmental factors should be included in studies on depressive symptoms (especially during the COVID-19 pandemic). In addition, integrating environmental factors allows for a better understanding of depressive symptoms, namely, to what extent environmental factors are risk and protective factors for depressive symptoms [27]. Furthermore, as some environmental factors such as sociodemographics (eg, employment rate and age distribution), economic conditions (eg, investments per employee), social affairs (eg, percentage of households with children), and living conditions (eg, percentage of 1-person households) can be passively measured via GPS data, the prediction and identification of depressive symptoms can be achieved with much lower costs, effort, and in a faster manner. However, it is unclear how accurately self-reported individual depressive symptoms can be predicted purely on the basis of GPS signals.

### Study Design

The aim of this study was to clarify to what extent environmental factors that can be passively measured via an individual's location tracked by GPS explained the statistical variance in individual depressive symptoms. In contrast to many studies that have focused on single predictors only [13], this study used a comprehensive approach and integrated a wide range of data available through GPS smartphone data (eg, sociodemographics, economic conditions, social affairs, and living conditions). Since this study was conducted during the COVID-19 pandemic, GPS-based COVID-19-specific variables such as the pandemic phase and the regional COVID-19 infection count were integrated as well.

The analyses were based on a data-driven statistical approach that allowed us to analyze a large number of predictor variables for individual depressive symptoms in an exploratory fashion without overfitting the data. Hence, there was no a priori selection of included variables based on any assumed connection to depressive symptoms, but all available variables of the data source were included.

## Methods

### Sample and Procedure

Data collection took place by means of the developed CORONA HEALTH app, a cooperative project between the German

universities of Würzburg, Ulm, and Regensburg, and the German Robert Koch Institute. The CORONA HEALTH app was specifically designed to explore factors during the COVID-19 pandemic and allows for both cross-sectional and longitudinal data analyses [28]; in this paper, we exclusively used cross-sectional data. Participants were recruited by means of public relation measures (eg, press releases and social networks). Participation in this study was voluntary, and participants did not receive incentives for participation. The inclusion criterion (for the analyzed study) was a minimum age of 18 years. In this study, only those data from smartphones with Android operating systems were analyzed. To participate, participants had to download the app and were then requested to provide informed consent (including use of GPS data). The study information provided was based on the quality standards set out by Beierle et al [29].

Based on the recorded GPS signal during survey participation, we assigned each participant to the nearest district, which was the smallest possible regional unit. The only GPS-based passive data collected were the participants' location at the time of the questionnaire. This location was used to link participants to district-level sociodemographic and environmental data from official German statistical sources. For data protection reasons, the resolution of the location was reduced to an accuracy of 11.1 km. In Germany, the statistical offices of the counties and the federal government provide current average values for 160 indicators at the regional level in the so-called Regional Atlas [30]. All these available average values for population structure, housing conditions, and so forth, were assigned to each participant according to his or her respective district. Data were gathered from July 2020 to February 2021, coinciding with the COVID-19 pandemic. This period in Germany was marked by several waves of infections and accompanying contact-restricting measures and correspondingly restricted public life [22]. Based on the date of the first questionnaire completion, the COVID-19 incidence rate in the respective district and information on the pandemic phase [22] were determined and assigned to the individual dataset.

The conception and implementation of the app followed medical device regulations and were certified accordingly (see the study by Holfelder et al [31]). This study was approved by the ethics committee of the University of Würzburg (130/20-me).

Due to the inclusion of exclusively passive data, the dataset did not contain any individual sociodemographic information but rather the sociodemographic averages in the respective district. The following sociodemographic averages describe the sample: those aged 0-17 years accounted for 15% of the population (averaged across all districts), those aged 18-24 years accounted for 8.3%, those aged 25-44 years accounted for 29.7%, those aged 45-64 years accounted for 25.6%, and those older than 65 years accounted for 27.5%. It is important to note that while those aged 0-17 years were mentioned as comprising 15% of the population, this figure provides general context about the population, as individuals younger than 18 years were not included in the study sample. On average, 39.6% of the population in the districts graduated from school with a general matriculation standard, and an average of 6% did not have a school-leaving certificate.

The original sample had 1760 participants. Eighteen individuals were dropped in a plausibility check based on correspondence between similar items, straightlining, intraindividual response variability, and extreme outliers. Of the remaining sample of participants, 60.4% (1052/1741) had an Android operating system, and 28.1% (490/1741) consented to the collection of passive data. Of the remaining sample, only those districts with more than 20 participants were included to meet the condition of normally distributed data in the clusters. This resulted in a final sample of 249 participants. Details on the sample can be found in Table S1 in [Multimedia Appendix 1](#).

## Measures

The primary outcome of this study was depressive symptoms, measured using the German version of the Patient Health Questionnaire-9 (PHQ-9) [32]. The PHQ-9 is a validated self-report instrument consisting of 9 items (eg, "Over the last week how often have you experienced little interest or pleasure in doing things?"), with responses on a 4-point scale (0="Not at all" to 3="Almost every day"). The total PHQ-9 score ranges from 0 to 27, with higher scores indicating more severe depressive symptoms. A score of 0-4 suggests minimal depressive symptoms; 5-9, mild symptoms; 10-14, moderate symptoms; 15-19, moderately severe symptoms; and 20-27, severe symptoms. The internal consistency of the PHQ-9 in this sample was excellent, with a Cronbach  $\alpha=0.85$ .

The PHQ-9 is widely used in both clinical and general populations and was chosen as the outcome measure because it allows for the assessment of the severity of depressive symptoms in a continuous format, which aligns with the study's objective to explore predictors of individual depressive symptoms. As a COVID-19 pandemic-specific variable, the number of infections per 100,000 inhabitants in the respective district at the time of assessment was included. These data were derived from official statistics reported by the Robert Koch Institute based on all laboratory-confirmed officially reported SARS-CoV-2 cases within the framework of the German Infection Protection Act [33]. In addition, the pandemic phase at the time of assessment was included in each case according to the classification by Schilling and colleagues [22] ranging from 1 (calendar weeks 10-20 in 2020) to 3 (calendar week 40 in 2020).

The respective regional environmental factors were determined using the district-level statistics of the Regional Atlas of the Federal Statistical Offices [30]. All variables available on the website on April 30, 2022, at the district level were included, with each construct included once. These included data on sociodemographics (eg, age group percentages in 2019, youth ratio in the regional population in 2021, percentage of the population without school-leaving qualifications in 2019, employment rate in 2020, disposable income per inhabitant in euros in 2020, and percentage of unemployed people in 2020), economy (eg, percentage of employed individuals by sector in 2019, investments per employee in thousand euros in 2019, and average length of stay of tourists in 2019), social affairs (percentage of households with children in 2011, percentage of children in care by age on March 1, 2021, proportion of fathers receiving parental benefits in 2014, and places in nursing homes

per 1000 inhabitants aged 65 years and older in 2020), and living environment (eg, percentage of 1-person households in 2011, population density of inhabitants per square kilometer in 2020, percentage of area by use in 2015, and consumption-based charge for drinking water supply per cubic meter in 2019). A complete list of variables is shown in Table S1 in [Multimedia Appendix 2](#). Due to the district clusters, the respective districts were included in the analyses as dummy variables. In addition, the pandemic phases were included as dummy variables. In total, the dataset included 121 variables.

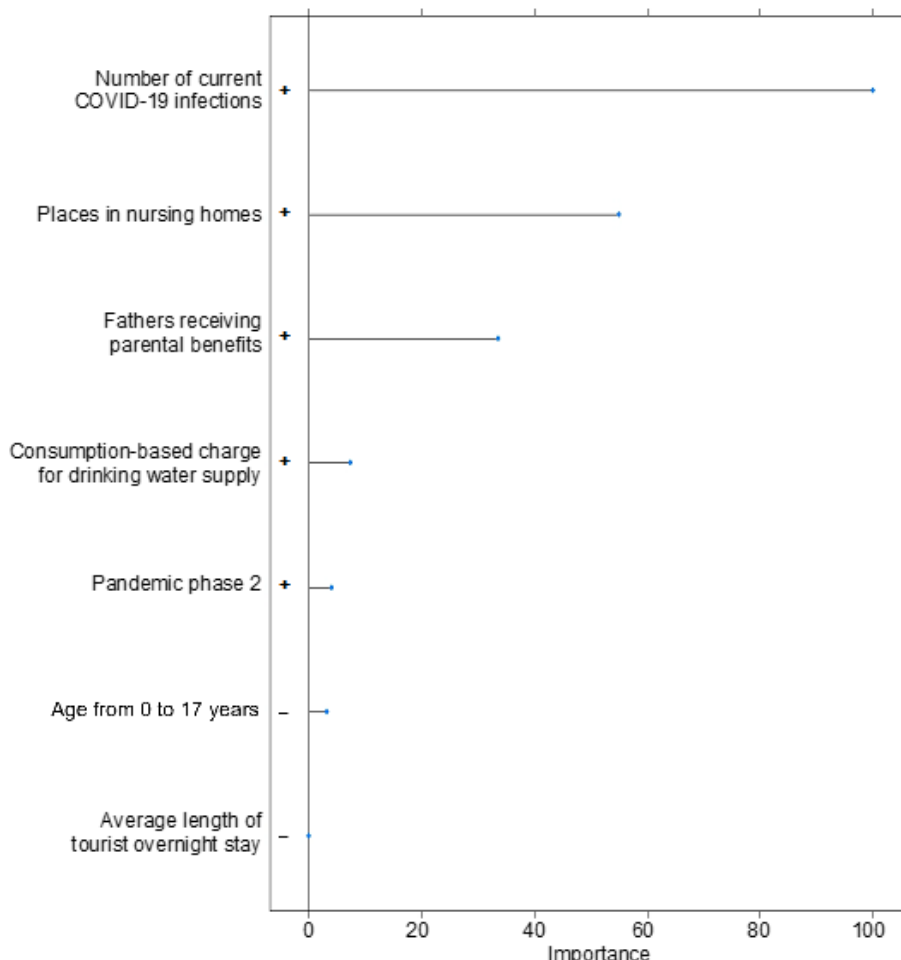
**Data Analyses**

Using GPS coordinates, each participant was assigned to a district. To control for the nested data, the respective districts were included in the analysis as dummy variables. Two participants who had a GPS location outside Germany were excluded. The data contained only 3 missing values (in the depressive symptoms variable); these were replaced using a single imputation with 500 iterations using classification and regression trees (cart method) with the R package “mice” (ie, the mice package for multivariate imputation by chained

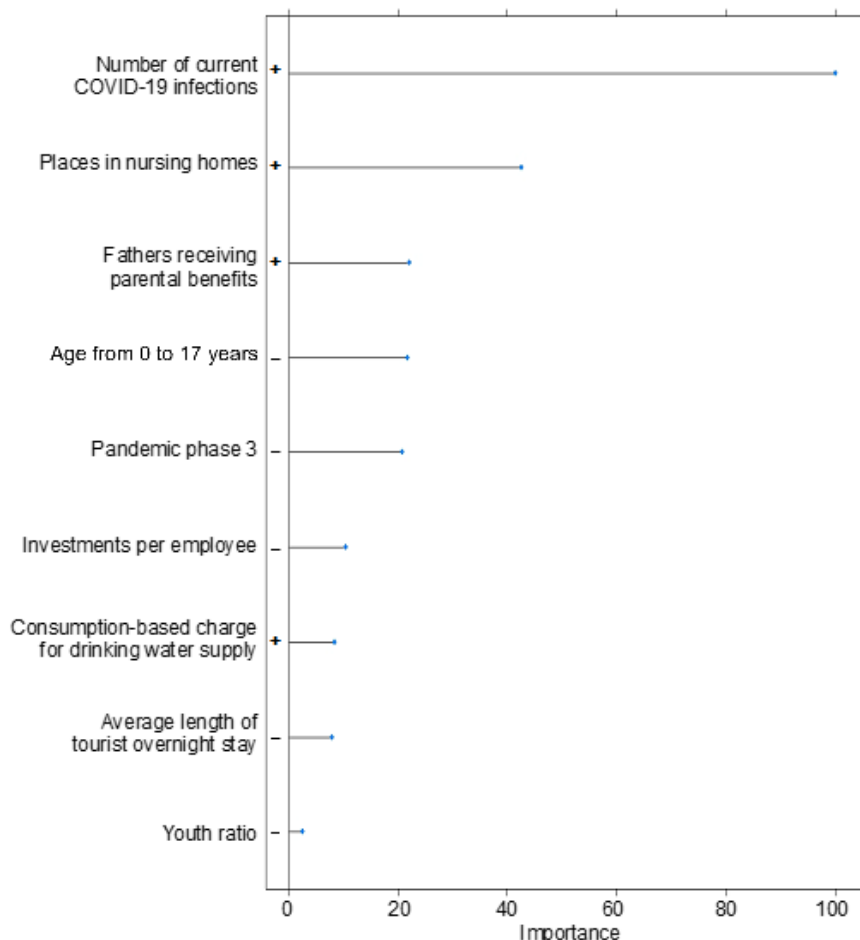
equations was developed by Stef van Buuren and is supported by Utrecht University) [34]. Metric variables were z standardized.

For the prediction of individual depressive symptoms, we chose to compare ridge, lasso, and elastic net regression models [35]. The predictive performance, signaling the best fit, relied on the root-mean-square error, mean absolute error, and explained variance ( $R^2$ ). Using a grid search, we implemented hyperparameter optimization. The analyses included automated variable selection. This allowed for a statistical check of which variables remained in the model with the best predictive performance, that is, the lowest degree of prediction errors [36]. To avoid overfitting, we split the dataset into a training (70%) and a test dataset (30%). In the training dataset, the final model was calculated using 10-fold cross-validation with 5-fold repetition [35]. In the test dataset, the model was evaluated for its predictive performance in a sample that had not been part of the model training. The result was a ranking by variable importance (Figures 1 and 2). Variables were ranked based on their contribution to the prediction of individual depressive symptoms.

**Figure 1.** Total variable importance as an indicator of the contribution to reduce the estimation error in the prediction of depressive symptoms in the lasso regression model.



**Figure 2.** Total variable importance as an indicator of the contribution to reduce the estimation error in the prediction of depressive symptoms in the elastic net regression model.



## Ethical Considerations

The conception and implementation of the app followed medical device regulations and were certified accordingly [31]. The original data collection for this study was approved by the ethics committee of the University of Würzburg (130/20-me). This approval explicitly covered both initial data collection and any potential future secondary analyses using the collected data. As such, no additional institutional review board approval or exemption was necessary for the present secondary analysis. In accordance with local regulations and institutional guidelines, the secondary use of existing, anonymized data does not require a separate review or approval. Specifically, the main institutional review board votum issued during the original data collection already included provisions for future secondary analyses. This is in line with German medical research standards, which state that when data are collected with informed consent that includes provisions for future use, and the data are anonymized, additional ethical approval is not required for secondary analyses (see Article 89 of the German Data Protection Regulation). All procedures adhered to institutional ethical standards for research involving human subjects, and all participant data were fully anonymized to protect privacy. Participation was voluntary, and no financial compensation was provided.

## Results

The comparison of the individual models (ridge, lasso, and elastic net regression) showed the following differences: in the prediction of individual depressive symptoms, lasso regression had the smallest average root-mean-square error (5.59). The explained variance was second to highest ( $R^2=0.05$ ). The elastic net regression produced the highest explained variance ( $R^2=0.06$ ; Table 1). The listed characteristic values showed lasso and elastic net regression as the analyses with the best fit. Lasso regression reduced the dataset from 121 variables to 7 variables, and elastic net regression reduced it from 121 to 9 variables.

The lasso and elastic net regression models shared the first 3 predictors. The most important predictor was the number of current COVID-19 infections in the respective district. The second was the number of places in nursing homes, and the third was the proportion of fathers receiving parental benefits. Both models also included the 3 variables population share of the 0-17 years age group, average length of stay of tourists, and the consumption-based charge for drinking water.

Lasso regression showed the second pandemic phase as a relevant predictor (Figure 1). Elastic net regression further listed (in descending variable importance) the third pandemic phase, investment per employee and the youth ratio as relevant predictors (Figure 2).

**Table 1.** Model fit indices for predicting depressive symptoms resulting from stepwise, ridge, lasso, and elastic net regression based on 50 replicate samples<sup>a</sup>.

	Minimum	1st quantile	Median	Mean	3rd quantile	Maximum
<b>MAE<sup>b</sup></b>						
Ridge	3.628	4.127	4.515	4.448	4.746	5.524
Lasso	3.606	4.231	4.440	4.464	4.643	5.717
Elastic net	3.605	4.162	4.506	4.483	4.741	5.764
<b>RMSE<sup>c</sup></b>						
Ridge	4.165	5.071	5.626	5.594	6.080	7.373
Lasso	4.216	5.244	5.630	5.585	5.942	7.104
Elastic net	4.169	5.219	5.677	5.618	5.912	7.162
<b>R<sup>2</sup></b>						
Ridge	0.000	0.003	0.019	0.048	0.079	0.228
Lasso	0.000	0.003	0.019	0.053	0.052	0.366
Elastic net	0.000	0.009	0.034	0.063	0.083	0.415

<sup>a</sup>Overall, elastic net regression had the best model fit due to the largest variance explained ( $R^2$ ), and lasso regression had the best model fit when including the smallest mean absolute error and root-mean-square error.

<sup>b</sup>MAE: mean absolute error.

<sup>c</sup>RMSE: root-mean-square error.

## Discussion

### Principal Findings

This study aimed to identify correlates of depressive symptoms based on regional factors allocated via GPS (passive data). It should be noted that the only GPS-based passive data collected were the participants' location at the time of the questionnaire, which was then used to assign participants to regional sociodemographic and environmental data based on their district. For this purpose, we used a passive data collection method. This had the advantage of integrating a wide range of data without any effort on the part of the participants.

The results helped identify regional and freely available variables as risk and protective factors for individual depressive symptoms. The identified risk and protective factors are presented as follows, starting with the 3 aspects that had the greatest relevance in both models.

The weekly number of cases of COVID-19 infections in the respective district was positively associated with higher individual depressive symptoms. In many cases, high COVID-19 case numbers were associated with contact restrictions in everyday life imposed by law or by the population itself [37]. A connection between these contact restrictions and individual depressive symptoms has been discussed repeatedly [38-41]. Different arguments were used to explain this connection. The reduction in social contact promoted loneliness, which has been identified as a risk factor for individual depressive symptoms [42,43]. In addition, an increase in misconduct and home violence has been reported [44]. Moreover, the high number of cases might have been associated with increased fear of infection, which increased substantially during the pandemic [41], a lack of care capacities, or a severe course of the disease

[45]. This could be related to individual depressive symptoms as an additional burden [46].

The proportion of fathers receiving a parental allowance was a risk factor for higher individual depressive symptoms. Parental allowance is a financial state benefit to mothers or fathers in Germany that is granted for a certain period during which the parent is not or only partially employed and devotes himself or herself to the care of their child. For methodological reasons, it was not possible to derive any causal associations from the results of our analysis. However, one possible explanation could be that paternal receipt of parental allowance might be an indicator for those families in which both parents are employed. These families were thus dependent on the functioning of the childcare infrastructure and therefore particularly challenged by the double burden of working from home and homeschooling for the duration of the pandemic. In a different study, increased individual depressive symptoms were also observed in parents of underage children during the COVID-19 pandemic, which also indicates the special burden on families [47].

In addition, the pandemic phase was a relevant factor for individual depressive symptoms. Phase 2 of the pandemic was positively related to individual depressive symptoms. The second phase of the pandemic, from the end of May to the end of September 2020, was accompanied by the opening of restaurants, gradual school openings, and the lifting of travel restrictions [22]. It is possible that with this return to regular daily life, depressive symptoms became more noticeable. In other words, it could be that it was only at the moment when a person was no longer able to cope with demands that he or she became aware of his or her own mental health condition. This could have been the case for those affected by depressive symptoms after the contact restrictions were removed and they

returned to everyday demands. The same reasoning could apply to the negative correlation between phase 3 of the pandemic and individual depressive symptoms. The third pandemic phase was characterized by a nationwide partial lockdown at the beginning of November 2021 with tightened contact restrictions and a nationwide lockdown with moderately strict regulations from mid-December and the Christmas holidays [22]. Due to contact restrictions and holidays, there were fewer day-to-day demands to attend to during the period, such as homeschooling, daily logistics due to travel, and social obligations. In the case of individual depressive symptoms associated with reduced drive and anhedonia [48], these individual depressive symptoms would have been less noticeable during the lockdown, as there were fewer events to manage anyway.

In contrast to previous findings, a high proportion of the 0-17 years age group and a high youth ratio were associated with lower levels of individual depressive symptoms. In contrast, numerous studies have shown that stress and individual depressive symptoms in families increased during the pandemic [47,49,50]. Since only proportions of the population and not the absolute numbers were collected, a large share of the 0-17 years age group can also mean that the proportion of families was correspondingly large and thus the proportion of people living alone correspondingly smaller. Research showed that 1-person households correlated with lower well-being compared with households with 2-4 people during the pandemic [51]. Furthermore, correlations between loneliness and individual depressive symptoms were also found [52].

In contrast to the present results, previous studies showed a correlation between nearby green spaces [25] or the quality of housing conditions (>60 m<sup>2</sup>, view from the windows, and poor indoor quality) [23] and lower depressive mood. A systematic review confirmed this and identified poor housing, lack of green spaces, and both noise and air pollution as environmental factors that were associated with depressive mood across studies [21]. The reason for the deviation of the present results could be that availability alone was not the decisive predictor for the association with depressive symptoms. Using the example of green spaces, research conducted during the COVID-19 pandemic, for example, showed that although natural areas were visited more often, visits to parks decreased [25]. The authors further discussed to what extent it was the time spent in green spaces that was related to low depressive symptoms or whether the green space was explicitly visited to participate in sports or to maintain social contacts and whether these were actually relevant predictors [25]. Sonnentag et al [53] found that the psychological mechanisms underlying recovery seem to be enhanced in natural environments.

### Strengths and Limitations

A central strength of this study is the application of a data-driven statistical approach that allowed us to analyze a large number

of predictor variables in an exploratory fashion without overfitting the data. By these means, the approach allowed us to investigate which environmental factors had predictive power. The collection of passive data via smartphone made it possible to reach participants despite pandemic-related contact restrictions and to make a corresponding statement about the included predictors for the COVID-19 pandemic period.

The cross-sectional study design of this study was limiting; statements about causal relationships were therefore not possible. In addition, the final dataset was very small, which might have weakened the significance of the analyses. A limitation that other studies have also observed [12] is the limited predictive power of individual depressive symptoms based on the passive data used. The use of GPS data was also a limitation, as it included only the location at the time of questionnaire processing. This was not the same as the place of residence of the respective person. How long the respective person was exposed to the general conditions could therefore not be inferred from the data. In addition, the informative value of the GPS data regarding individual environmental conditions was limited due to the regional resolution: since large variances in the average values were possible, it was of course not possible to draw conclusions about individual cases. Moreover, the image of the environment was only very roughly described. However, it represented the smallest possible regional size at which data from statistical offices could be collected nationwide. In addition, a higher spatial resolution, for example, per block of houses, would make it impossible to anonymize the data, as too few people would remain in the respective cluster. The validity of this study is further limited by the nonrepresentative dataset. As this analysis was limited to data from Android smartphones and required agreement to the collection of passive data, selection bias could not be ruled out. This must be taken into account when evaluating the results of this study. In addition, neither changes in predictors over time nor long-term data collection was considered to depict cause and effect more validly. This study can therefore make statements only about correlations and not about cause and effect.

### Conclusions

This study showed that regional average data on socioeconomics and living environment seemed to be limited predictors of individual depressive symptoms; for the pandemic period, individual depressive symptoms were predicted, although the predictors had a rather low explanatory value and are not transferable to postpandemic periods. This suggests that the spatial resolution would need to be modified or that the dataset would need to be supplemented with additional data sources. Nevertheless, the results help classify freely available average data as risk or protective factors for individual depressive symptoms.

### Acknowledgments

We thank Johannes Allgaier, Fabian Haug, Julian Haug, and Carsten Vogel (University of Würzburg); Marc Holfelder (LA2 GmbH); Klaus Kammerer and Johannes Schobel (University of Ulm); Winfried Schlee (University of Regensburg); and Michael Schultz (typelover) for their support and contributions to the implementation of and feedback on the project.

## Authors' Contributions

JSE and MW developed the research questions and methodology in consultation with CC. MW and HS advised the planning of the statistical analyses. RP developed the app and data collection processes. JSE and MW performed the data analyses and wrote the manuscript in consultation with HS, RP, HB, and CC. All authors critically revised the paper and approved the final revision to be published.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Descriptive statistics of the present sample of 249 German-speaking adults across districts.

[\[DOCX File, 23 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

A list of variables from the Regional Atlas of Germany detailing sociodemographic, economic, and environmental data used in the study.

[\[DOCX File, 52 KB-Multimedia Appendix 2\]](#)

## References

1. Lépine JP, Briley M. The increasing burden of depression. *Neuropsychiatr Dis Treat*. 2011;7(1):3-7. [doi: [10.2147/NDT.S19617](https://doi.org/10.2147/NDT.S19617)] [Medline: [21750622](https://pubmed.ncbi.nlm.nih.gov/21750622/)]
2. Wahlbeck K. Public mental health: the time is ripe for translation of evidence into practice. *World Psychiatry*. 2015;14(1):36-42. [FREE Full text] [doi: [10.1002/wps.20178](https://doi.org/10.1002/wps.20178)] [Medline: [25655149](https://pubmed.ncbi.nlm.nih.gov/25655149/)]
3. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Med*. 2015;13(1):72. [FREE Full text] [doi: [10.1186/s12916-015-0325-4](https://doi.org/10.1186/s12916-015-0325-4)] [Medline: [25879936](https://pubmed.ncbi.nlm.nih.gov/25879936/)]
4. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*. 2016;41(7):1691-1696. [FREE Full text] [doi: [10.1038/npp.2016.7](https://doi.org/10.1038/npp.2016.7)] [Medline: [26818126](https://pubmed.ncbi.nlm.nih.gov/26818126/)]
5. Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Transl Psychiatry*. 2017;7(3):e1053. [FREE Full text] [doi: [10.1038/tp.2017.25](https://doi.org/10.1038/tp.2017.25)] [Medline: [28267146](https://pubmed.ncbi.nlm.nih.gov/28267146/)]
6. He X. Depression diagnosis and forecast based on mobile phone sensor data. 2022. Presented at: 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2022 July 11-15:4679-4682; Glasgow, Scotland, United Kingdom. [doi: [10.1109/embc48229.2022.9871255](https://doi.org/10.1109/embc48229.2022.9871255)]
7. Seppälä J, De Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review. *JMIR Ment Health*. 2019;6(2):e9819. [FREE Full text] [doi: [10.2196/mental.9819](https://doi.org/10.2196/mental.9819)] [Medline: [30785404](https://pubmed.ncbi.nlm.nih.gov/30785404/)]
8. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR Mhealth Uhealth*. 2021;9(7):e26540. [doi: [10.2196/26540](https://doi.org/10.2196/26540)] [Medline: [34255713](https://pubmed.ncbi.nlm.nih.gov/34255713/)]
9. Shakeri Hossein Abad Z, Kline A, Sultana M, Noaen M, Nurmambetova E, Lucini F, et al. Digital public health surveillance: a systematic scoping review. *NPJ Digit Med*. 2021;4(1):41. [FREE Full text] [doi: [10.1038/s41746-021-00407-6](https://doi.org/10.1038/s41746-021-00407-6)] [Medline: [33658681](https://pubmed.ncbi.nlm.nih.gov/33658681/)]
10. Birk RH, Samuel G. Digital phenotyping for mental health: reviewing the challenges of using data to monitor and predict mental health problems. *Curr Psychiatry Rep*. 2022;24(10):523-528. [doi: [10.1007/s11920-022-01358-9](https://doi.org/10.1007/s11920-022-01358-9)] [Medline: [36001220](https://pubmed.ncbi.nlm.nih.gov/36001220/)]
11. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J Med Internet Res*. 2017;19(3):e75. [FREE Full text] [doi: [10.2196/jmir.6678](https://doi.org/10.2196/jmir.6678)] [Medline: [28302595](https://pubmed.ncbi.nlm.nih.gov/28302595/)]
12. Currey D, Torous J. Digital phenotyping correlations in larger mental health samples: analysis and replication. *BJPsych Open*. 2022;8(4):e106. [FREE Full text] [doi: [10.1192/bjo.2022.507](https://doi.org/10.1192/bjo.2022.507)] [Medline: [35657687](https://pubmed.ncbi.nlm.nih.gov/35657687/)]
13. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. *J Med Internet Res*. 2017;19(7):e262. [FREE Full text] [doi: [10.2196/jmir.7006](https://doi.org/10.2196/jmir.7006)] [Medline: [28739561](https://pubmed.ncbi.nlm.nih.gov/28739561/)]
14. Opoku Asare K, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: a longitudinal data analysis. *Pervasive Mobile Comput*. 2022;83:101621. [doi: [10.1016/j.pmcj.2022.101621](https://doi.org/10.1016/j.pmcj.2022.101621)]



15. Wang R, Wang W, da Silva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proc ACM Interactive Mobile Wearable Ubiquitous Technol.* 2018;2(1):1-26. [doi: [10.1145/3191775](https://doi.org/10.1145/3191775)]
16. Liu D, Feng XL, Ahmed F, Shahid M, Guo J. Detecting and measuring depression on social media using a machine learning approach: systematic review. *JMIR Ment Health.* 2022;9(3):e27244. [FREE Full text] [doi: [10.2196/27244](https://doi.org/10.2196/27244)] [Medline: [35230252](https://pubmed.ncbi.nlm.nih.gov/35230252/)]
17. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. *Front Psychiatry.* 2021;12(43):625247. [FREE Full text] [doi: [10.3389/fpsy.2021.625247](https://doi.org/10.3389/fpsy.2021.625247)] [Medline: [33584388](https://pubmed.ncbi.nlm.nih.gov/33584388/)]
18. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. *JMIR Mhealth Uhealth.* 2018;6(8):e165. [FREE Full text] [doi: [10.2196/mhealth.9691](https://doi.org/10.2196/mhealth.9691)] [Medline: [30104184](https://pubmed.ncbi.nlm.nih.gov/30104184/)]
19. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. The relationship between clinical, momentary, and sensor-based assessment of depression. *Int Conf Pervasive Comput Technol Healthc.* 2015;2015:229-232. [FREE Full text] [doi: [10.4108/icst.pervasivehealth.2015.259034](https://doi.org/10.4108/icst.pervasivehealth.2015.259034)] [Medline: [26640739](https://pubmed.ncbi.nlm.nih.gov/26640739/)]
20. Sükei E, Norbury A, Perez-Rodriguez MM, Olmos PM, Artés A. Predicting emotional states using behavioral markers derived from passively sensed data: data-driven machine learning approach. *JMIR Mhealth Uhealth.* 2021;9(3):e24465. [FREE Full text] [doi: [10.2196/24465](https://doi.org/10.2196/24465)] [Medline: [33749612](https://pubmed.ncbi.nlm.nih.gov/33749612/)]
21. Rautio N, Filatova S, Lehtiniemi H, Miettunen J. Living environment and its relationship to depressive mood: a systematic review. *Int J Soc Psychiatry.* 2018;64(1):92-103. [doi: [10.1177/0020764017744582](https://doi.org/10.1177/0020764017744582)] [Medline: [29212385](https://pubmed.ncbi.nlm.nih.gov/29212385/)]
22. Schilling J, Buda S, Fischer M. Retrospektive Phaseneinteilung der COVID-19-Pandemie in Deutschland bis Februar. Germany. Robert Koch-Institut; 2021:3-12.
23. Amerio A, Brambilla A, Morganti A, Aguglia A, Bianchi D, Santi F, et al. COVID-19 lockdown: housing built environment's effects on mental health. *Int J Environ Res Public Health.* 2020;17(16):5973. [FREE Full text] [doi: [10.3390/ijerph17165973](https://doi.org/10.3390/ijerph17165973)] [Medline: [32824594](https://pubmed.ncbi.nlm.nih.gov/32824594/)]
24. Groot J, Keller A, Joensen A, Nguyen TL, Nybo Andersen AM, Strandberg-Larsen K. Impact of housing conditions on changes in youth's mental health following the initial national COVID-19 lockdown: a cohort study. *Sci Rep.* 2022;12(1):1939. [doi: [10.1038/s41598-022-04909-5](https://doi.org/10.1038/s41598-022-04909-5)] [Medline: [35121742](https://pubmed.ncbi.nlm.nih.gov/35121742/)]
25. Löhmus M, Stenfors CUD, Lind T, Lauber A, Georgelis A. Mental health, greenness, and nature related behaviors in the adult population of stockholm county during COVID-19-related restrictions. *Int J Environ Res Public Health.* 2021;18(6):3303. [FREE Full text] [doi: [10.3390/ijerph18063303](https://doi.org/10.3390/ijerph18063303)] [Medline: [33806787](https://pubmed.ncbi.nlm.nih.gov/33806787/)]
26. Ramiz L, Contrand B, Rojas Castro MY, Dupuy M, Lu L, Sztal-Kutas C, et al. A longitudinal study of mental health before and during COVID-19 lockdown in the French population. *Global Health.* 2021;17(1):29. [doi: [10.1186/s12992-021-00682-8](https://doi.org/10.1186/s12992-021-00682-8)] [Medline: [33752717](https://pubmed.ncbi.nlm.nih.gov/33752717/)]
27. Han J, Zhang Z, Mascolo C, Andre E, Tao J, Zhao Z, et al. Deep learning for mobile mental health: challenges and recent advances. *IEEE Signal Process Mag.* 2021;38(6):96-105. [doi: [10.1109/MSP.2021.3099293](https://doi.org/10.1109/MSP.2021.3099293)]
28. Beierle F, Schobel J, Vogel C, Allgaier J, Mulansky L, Haug F, et al. Corona health—a study- and sensor-based mobile app platform exploring aspects of the COVID-19 pandemic. *Int J Environ Res Public Health.* 2021;18(14):7395. [FREE Full text] [doi: [10.3390/ijerph18147395](https://doi.org/10.3390/ijerph18147395)] [Medline: [34299846](https://pubmed.ncbi.nlm.nih.gov/34299846/)]
29. Beierle F, Tran VT, Allemand M, Neff P, Schlee W, Probst T, et al. Context data categories and privacy model for mobile data collection apps. *Procedia Comput Sci.* 2018;134:18-25. [doi: [10.1016/j.procs.2018.07.139](https://doi.org/10.1016/j.procs.2018.07.139)]
30. Statistical Offices of the Federal and State Governments. Joint statistics portal. URL: <https://regionalatlas.statistikportal.de/> [accessed 2022-04-30]
31. Holfelder M, Mulansky L, Schlee W, Baumeister H, Schobel J, Greger H, et al. Medical device regulation efforts for mHealth apps during the COVID-19 pandemic—an experience report of corona check and corona health. *J Multidisciplinary Sci J.* 2021;4(2):206-222. [doi: [10.3390/j4020017](https://doi.org/10.3390/j4020017)]
32. Löwe B. PHQ-D manual. Komplettversion und Kurzform. 2002. URL: [https://www.klinikum.uni-heidelberg.de/fileadmin/Psychosomatische\\_Klinik/download/PHQ\\_Kurzanleitung1.pdf](https://www.klinikum.uni-heidelberg.de/fileadmin/Psychosomatische_Klinik/download/PHQ_Kurzanleitung1.pdf) [accessed 2024-10-14]
33. Federal Ministry of Justice. Act on the prevention and control of infectious diseases in humans. URL: <https://www.gesetze-im-internet.de/ifsg/index.html> [accessed 2000-07-01]
34. van Buuren S. Multivariate imputation by chained equations. URL: <https://cran.r-project.org/web/packages/mice/mice.pdf> [accessed 2021-01-09]
35. James G. *An Introduction to Statistical Learning With Applications in R.* New York. Springer; 2013.
36. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B.* 2005;67(2):301-320. [doi: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x)]
37. Borkowski P, Jażdżewska-Gutta M, Szmelter-Jarosz A. Lockdowned: everyday mobility changes in response to COVID-19. *J Transp Geogr.* 2021;90:102906. [FREE Full text] [doi: [10.1016/j.jtrangeo.2020.102906](https://doi.org/10.1016/j.jtrangeo.2020.102906)] [Medline: [35721765](https://pubmed.ncbi.nlm.nih.gov/35721765/)]

38. Benke C, Autenrieth LK, Asselmann E, Pané-Farré CA. Lockdown, quarantine measures, and social distancing: associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany. *Psychiatry Res.* 2020;293:113462. [[FREE Full text](#)] [doi: [10.1016/j.psychres.2020.113462](https://doi.org/10.1016/j.psychres.2020.113462)] [Medline: [32987222](#)]
39. Bonati M, Campi R, Segre G. Psychological impact of the quarantine during the COVID-19 pandemic on the general European adult population: a systematic review of the evidence. *Epidemiol Psychiatr Sci.* 2022;31:e27. [[FREE Full text](#)] [doi: [10.1017/S2045796022000051](https://doi.org/10.1017/S2045796022000051)] [Medline: [35475479](#)]
40. Mata J, Wenz A, Rettig T, Reifenscheid M, Möhring K, Krieger U, et al. Health behaviors and mental health during the COVID-19 pandemic: a longitudinal population-based survey in Germany. *Soc Sci Med.* 2021;287:114333. [[FREE Full text](#)] [doi: [10.1016/j.socscimed.2021.114333](https://doi.org/10.1016/j.socscimed.2021.114333)] [Medline: [34455337](#)]
41. Petzold MB, Bendau A, Plag J, Pyrkosch L, Mascarell Maricic L, Betzler F, et al. Risk, resilience, psychological distress, and anxiety at the beginning of the COVID-19 pandemic in Germany. *Brain Behav.* 2020;10(9):e01745. [[FREE Full text](#)] [doi: [10.1002/brb3.1745](https://doi.org/10.1002/brb3.1745)] [Medline: [32633464](#)]
42. Hoffart A, Johnson SU, Ebrahimi OV. Loneliness and social distancing during the COVID-19 pandemic: risk factors and associations with psychopathology. *Front Psychiatry.* 2020;11:589127. [[FREE Full text](#)] [doi: [10.3389/fpsy.2020.589127](https://doi.org/10.3389/fpsy.2020.589127)] [Medline: [33329136](#)]
43. Palgi Y, Shrira A, Ring L, Bodner E, Avidor S, Bergman Y, et al. The loneliness pandemic: loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. *J Affect Disord.* 2020;275:109-111. [[FREE Full text](#)] [doi: [10.1016/j.jad.2020.06.036](https://doi.org/10.1016/j.jad.2020.06.036)] [Medline: [32658811](#)]
44. Thiel F, Büechl VCS, Rehberg F, Mojahed A, Daniels JK, Schellong J, et al. Changes in prevalence and severity of domestic violence during the COVID-19 pandemic: a systematic review. *Front Psychiatry.* 2022;13:874183. [[FREE Full text](#)] [doi: [10.3389/fpsy.2022.874183](https://doi.org/10.3389/fpsy.2022.874183)] [Medline: [35492711](#)]
45. Schmitz A, Garten C, Kühne S, Brandt M. Worries about inadequate medical treatment in case of a COVID-19 infection: the role of social inequalities, COVID-19 prevalence and healthcare infrastructure. *BMC Public Health.* 2022;22(1):1761. [[FREE Full text](#)] [doi: [10.1186/s12889-022-14024-9](https://doi.org/10.1186/s12889-022-14024-9)] [Medline: [36114486](#)]
46. Ingram RE, Luxton DD. Vulnerability-stress models. In: Hankin BL, Abela JRZ, editors. *Development of Psychopathology: A Vulnerability-Stress Perspective*. Thousand Oaks, CA. Sage; 2005:32-46.
47. Johnson M, Skjerdingstad N, Ebrahimi OV, Hoffart U, Johnson SU. Parenting in a pandemic: parental stress, anxiety and depression among parents during the government-initiated physical distancing measures following the first wave of COVID-19. *Stress Health.* 2022;38(4):637-652. [doi: [10.1002/smi.3120](https://doi.org/10.1002/smi.3120)] [Medline: [34902219](#)]
48. World Health Organization. *Pocket Guide to the ICD-10 Classification of Mental Disorders*. Germany. Hogrefe; 2019.
49. Calvano C, Engelke L, Di Bella J, Kindermann J, Renneberg B, Winter SM. Families in the COVID-19 pandemic: parental stress, parent mental health and the occurrence of adverse childhood experiences-results of a representative survey in Germany. *Eur Child Adolesc Psychiatry.* 2022;31(7):1-13. [[FREE Full text](#)] [doi: [10.1007/s00787-021-01739-0](https://doi.org/10.1007/s00787-021-01739-0)] [Medline: [33646416](#)]
50. Adams EL, Smith D, Caccavale LJ, Bean MK. Parents are stressed! Patterns of parent stress across COVID-19. *Front Psychiatry.* 2021;12:626456. [[FREE Full text](#)] [doi: [10.3389/fpsy.2021.626456](https://doi.org/10.3389/fpsy.2021.626456)] [Medline: [33897489](#)]
51. Chen D, Wang Y. Inequality-related health and social factors and their impact on well-being during the COVID-19 pandemic: findings from a national survey in the UK. *Int J Environ Res Public Health.* 2021;18(3):1014. [[FREE Full text](#)] [doi: [10.3390/ijerph18031014](https://doi.org/10.3390/ijerph18031014)] [Medline: [33498842](#)]
52. Werner AM, Tibubos AN, Mülder LM, Reichel JL, Schäfer M, Heller S, et al. The impact of lockdown stress and loneliness during the COVID-19 pandemic on mental health among university students in Germany. *Sci Rep.* 2021;11(1):22637. [[FREE Full text](#)] [doi: [10.1038/s41598-021-02024-5](https://doi.org/10.1038/s41598-021-02024-5)] [Medline: [34811422](#)]
53. Sonnentag S, Venz L, Casper A. Advances in recovery research: what have we learned? What should be done next? *J Occup Health Psychol.* 2017;22(3):365-380. [doi: [10.1037/ocp0000079](https://doi.org/10.1037/ocp0000079)] [Medline: [28358572](#)]

---

## Abbreviations

**PHQ-9:** Patient Health Questionnaire-9

---

*Edited by T de Azevedo Cardoso; submitted 30.09.23; peer-reviewed by M Hamlin, B Davis; comments to author 14.05.24; revised version received 22.08.24; accepted 24.09.24; published 03.12.24*

*Please cite as:*

*Edler J-S, Winter M, Steinmetz H, Cohrdes C, Baumeister H, Pryss R*

*Predicting Depressive Symptoms Using GPS-Based Regional Data in Germany With the CORONA HEALTH App During the COVID-19 Pandemic: Cross-Sectional Study*

*Interact J Med Res 2024;13:e53248*

*URL: <https://www.i-jmr.org/2024/1/e53248>*

*doi: [10.2196/53248](https://doi.org/10.2196/53248)*

*PMID: [39625745](https://pubmed.ncbi.nlm.nih.gov/39625745/)*

©Johanna-Sophie Edler, Michael Winter, Holger Steinmetz, Caroline Cohrdes, Harald Baumeister, Rüdiger Pryss. Originally published in the Interactive Journal of Medical Research (<https://www.i-jmr.org/>), 03.12.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.i-jmr.org/>, as well as this copyright and license information must be included.