

Original Paper

Statistical Methods for Item Reduction in a Representative Lifestyle Questionnaire: Pilot Questionnaire Study

Alessio Staffini^{1,2,3}, MSc; Kento Fujita⁴, MSc; Akiko Kishi Svensson^{1,5,6}, MD, MPH, PhD; Ung-Il Chung^{1,7,8}, MD, PhD; Thomas Svensson^{1,5,7}, MD, PhD

¹Precision Health, Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

²Department of Economics and Finance, Catholic University of Milan, Milan, Italy

³Data Solutions Division, Project Promotion Department, Albert Inc, Tokyo, Japan

⁴Data Service Infrastructure Development Department, Service Infrastructure Division, IT-OT Innovation Division, Mobile Technology Unit, SoftBank Corp, Tokyo, Japan

⁵Department of Clinical Sciences, Lund University, Malmö, Sweden

⁶Department of Diabetes and Metabolic Diseases, The University of Tokyo, Tokyo, Japan

⁷School of Health Innovation, Kanagawa University of Human Services, Kawasaki, Japan

⁸Clinical Biotechnology, Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding Author:

Akiko Kishi Svensson, MD, MPH, PhD

Precision Health, Department of Bioengineering

Graduate School of Engineering

The University of Tokyo

7-3-1 Hongo, Bunkyo-ku

Tokyo, 113-8656

Japan

Phone: 81 358414737

Email: kishi@bioeng.t.u-tokyo.ac.jp

Abstract

Background: Reducing the number of items in a questionnaire while maintaining relevant information is important as it is associated with advantages such as higher respondent engagement and reduced response error. However, in health care, after the original design, an *a posteriori* check of the included items in a questionnaire is often overlooked or considered to be of minor importance. When conducted, this is often based on a single selected method. We argue that before finalizing any lifestyle questionnaire, a *a posteriori* validation should always be conducted using multiple approaches to ensure the robustness of the results.

Objective: The objectives of this study are to compare the results of two statistical methods for item reduction (variance inflation factor [VIF] and factor analysis [FA]) in a lifestyle questionnaire constructed by combining items from different sources and analyze the different results obtained from the 2 methods and the conclusions that can be made about the original items.

Methods: Data were collected from 79 participants (heterogeneous in age and sex) with a high risk of metabolic syndrome working in a financial company based in Tokyo. The lifestyle questionnaire was constructed by combining items (asked with daily, weekly, and monthly frequency) from multiple validated questionnaires and other selected questions. Item reduction was conducted using VIF and exploratory FA. Adequacy tests were used to check the data distribution and sampling adequacy.

Results: Among the daily and weekly questions, both VIF and FA identified redundancies in sleep-related items. Among the monthly questions, both approaches identified redundancies in stress-related items. However, the number of items suggested for reduction often differed: VIF suggested larger reductions than FA for daily questions but fewer reductions for weekly questions. Adequacy tests always confirmed that the structural detection was adequate for the considered items.

Conclusions: As expected, our analyses showed that VIF and FA produced both similar and different findings, suggesting that questionnaire designers should consider using multiple methods for item reduction. Our findings using both methods indicate that many questions, especially those related to sleep, are redundant, indicating that the considered lifestyle questionnaire can be shortened.

KEYWORDS

item reduction; surveys and lifestyle questionnaires; feedback measures; questionnaire design; variance inflation factor; factor analysis; mobile phone

Introduction

Background

Short questionnaires have several advantages over long ones: they are less expensive to design [1]; are associated with less random or systematic error or noise in the reported results caused by lack of motivation, fatigue, or boredom [2-4]; and have higher respondent engagement, with empirical evidence indicating that long questionnaires are associated with a low response rate [5]. Therefore, reducing the number of survey items is important.

A number of approaches to reduce survey items have been studied. First, in psychometrics, the Cronbach α coefficient is often used in Classical Test Theory [6] to assess the internal consistency of a given questionnaire. Second, factor analysis (FA) and principal component analysis were used to extract a latent structure and explain most of the data variance with fewer factors or components than the original questionnaire items (eg, the studies by McHorney et al [7], Bai et al [8], and Brosnan et al [9]). Third, researchers have also focused on Item Response Theory (IRT) [10,11]; however, IRT methods are most effective when the original questionnaire or survey is developed using IRT or when there are theoretical reasons to expect it to fit an IRT model [12]. Recent studies have used the variance inflation factor (VIF) to better address collinearity problems among covariates when performing regression analysis of survey data [13]. However, despite the advantages and practicality of automatic variable selection, simulations suggest that the VIF does not necessarily identify a true underlying model [14]. In general, it is better not to rely on a single method for variable reduction and instead to compare the results of multiple approaches.

Although it is important to minimize errors in questionnaire responses in all fields, it is paramount in medicine and public health, where omissions and inaccuracies can lead to possible misdiagnoses and subsequent incorrect treatments or interventions. Despite this, emphasis in health care is often placed only on the initial design, where a questionnaire is typically evaluated by comparing its internal consistency with that of other similar questionnaires or with a different version of the same questionnaire [15-17]. The questionnaires also sometimes include open-ended questions, despite evidence showing that such questions often do not provide sufficiently solid insights [18]. Although an *a priori* carefully designed questionnaire is extremely important, we believe that including an *a posteriori* check of the questions before making any result-based inferences will help improve the overall quality of the instrument and should, therefore, become part of standard procedure.

Objective

Our study builds upon other research conducted in this direction: to name a few related works, Cappelleri et al [19] applied FA to develop a questionnaire to measure the satisfaction of patients with type 1 diabetes, highlighting that the two key factors were convenience and social comfort; Juniper et al [20] compared the results of the impact method, which preserves items according to their relative importance as perceived by the patients, with the results of FA in a quality of life questionnaire, suggesting that different approaches lead to different results; Arifin and Yusoff [21] applied confirmatory FA, as well as performed other statistical tests, for an emotional intelligence inventory to be used among medical course applicants. In the Classical Test Theory branch, researchers often compared Cronbach α with Rasch analysis; for example, Prieto et al [22] did so on a 38-item health questionnaire, concluding that the methods led to similar results; Erhart et al [23] analyzed the possibility of performing item reduction by comparing the results of Cronbach α with those of Rasch item fit for a health-related quality of life questionnaire administered to children and adolescents, concluding, on the other hand, that both methods should be accompanied by additional analyses.

We found that very often, researchers focused on a single technique or similar approaches without comparing the results from multiple different methods. This was the starting point of our study. Despite its desirable properties (such as automatic identification and removal of multicollinear items) and statistical foundations, VIF is rarely directly used as an item reduction technique. FA indirectly deals with the same multicollinearity problem by grouping *similar* variables under the same factor; however, its inner logic is different, as multicollinearity is solved because of dimensionality reduction. To the best of our knowledge, this is the first study to directly compare how similar or how different the results obtained from these methods are when applied to the items of a lifestyle questionnaire.

The aim of this study is to compare the results of VIF and FA for item reduction in a lifestyle questionnaire constructed by combining items from different sources, with questions asked with daily, weekly, or monthly frequency.

Methods

Study Participants

The collected questionnaire data (described in detail in the following sections) form part of a 1-month pilot study for lifestyle interventions, with a more general aim than the one described in this paper; that is, in particular, to investigate the feasibility of using a smartphone lifestyle intervention app and assess the feasibility of collecting data from a wearable device. The recruited participants were individuals with a high risk of metabolic syndrome. The sample size was calculated based on

accepting an α risk of .05 and a β risk of .25 in a 1-tailed test, with an estimated 10% loss to follow-up. The estimated sample size was 120 participants. A total of 117 individuals consented to participate; however, 2 (1.7%) individuals were withdrawn, leaving a total of 115 (98.3%) participants. The 115 participants were randomly allocated into 2 groups: the intervention group (79/115, 68.7% individuals who followed a lifestyle education program over the study period) and the control group (36/115, 31.3% individuals). Allocation to the intervention and control groups occurred through randomization, with stratification by sex and age (<40 years or >40 years). Participants in the intervention group (the focus of this study), men and women aged between 29 and 58 years, were selected from a financial company based in Tokyo. To minimize sampling bias in the collected questionnaire data, we confirmed that the intervention group was sufficiently heterogeneous in terms of age, sex, and employment conditions. The initial study excluded those who had a history of serious medical conditions, had received any other lifestyle intervention, planned to take long vacations, had night shifts, and were pregnant (or those with suspected pregnancy). The study period of the research presented in this paper was from March 2 to March 30, 2018.

All research was performed in accordance with relevant ethical guidelines and regulations. All the participants received detailed information about the purpose of the study in writing and during explanatory face-to-face meetings. All participants provided written informed consent and understood that participation was completely voluntary and could be discontinued at any time without any disadvantage or penalty. Participants were given a wearable device as an incentive for participation.

Questionnaire

The intervention group participants were asked to respond to lifestyle-related questions through a smartphone-based mobile app for a month. For the purposes of this study, we included participants who answered any of the questions.

We constructed a lifestyle questionnaire comprising 51 questions, focusing in particular on the domains of sleep, stress, nutrition, and alcohol and tobacco consumption. For brevity, the questionnaire items are indicated by *item X*, where *X* represents the item number. The questionnaire is provided in [Multimedia Appendix 1](#).

Daily questions on sleep quality were selected from the validated St Mary's Hospital Sleep Questionnaire (items 2-5) [24]. Daily questions to assess dietary habits were taken from the Dietary Guidelines provided by the Ministry of Agriculture, Forestry, and Fisheries of Japan (Food Safety and Consumer Affairs Bureau; items 6-16). Daily questions about caffeine intake (items 17 and 18) and stress levels (items 20 and 21) have been widely used in many population-based cohort studies. In addition, we added 2 other daily questions, not taken from any validated questionnaire, about opening the smartphone-based app in the morning (item 1) and alcohol consumption (item 19). Most of the weekly questions (21/26, 81%) were taken from the validated General Sleep Disturbance Scale (items 25-45) [25]. We added 5 other questions to assess the participants' commitment to reducing alcohol intake (item 47), determine the number of cigarettes smoked per day over the past week and interest in a

smoking cessation program (items 23 and 46), and information about working conditions (items 22 and 24). Finally, monthly questions were taken from the validated Perceived Stress Scale-4 (items 48-51) [26]. As shown in [Multimedia Appendix 1](#), items 48 and 51 are reverse-coded items; this is sometimes recommended to cross-check the validity of the responses.

When a participant did not complete a daily questionnaire item, the data were treated as missing and excluded from the analysis. However, when a participant did not complete a weekly or monthly questionnaire item, we assumed that their subsequent response to the item was applicable for the entire respective week or month and, thus, backfilled missing values.

Statistical Analyses

Overview

All statistical analyses were performed using Python (version 3.7.4) on the Anaconda platform (Anaconda Inc). To perform effective item reduction, it is important to verify that the variables we want to exclude, in fact, explain the same underlying variability as the other variables remaining in the questionnaire. We considered two statistical approaches: VIF and FA.

VIF Analysis

VIF is the quotient of the variance in a model with multiple variables and the variance in a model with only 1 variable. It indicates the strength of multicollinearity among a set of variables, assuming that they have a linear relationship. Each variable, in turn, is regressed on all other variables present in the set. Considering a set *A* of *n* variables, the VIF associated with variable $i \in A$ is defined as follows:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Here, R_i^2 is the coefficient of determination obtained by ordinary least squares regression and regressing variable *i* on all the other $j \neq i$ variables in the set. A higher VIF_i reports greater collinearity between variable *i* and the other predictors. As $0 \leq R_i^2 \leq 1$, $VIF_i \in [1, \infty)$.

There is no general consensus on the ideal VIF threshold for indicating multicollinearity, with many different values having been used in the empirical literature. For example, Vittinghoff et al [27] suggested a threshold value of 10, whereas Johnston et al [28] were more conservative and used a threshold value of 2.5. Keeping in mind that multicollinearity is a bigger problem with a small sample size [29], we selected a threshold value of 5; this cutoff is often selected to establish a high risk of multicollinearity [30,31].

We applied VIF iteratively: first, we calculated VIF_i for each variable in set *A*; then, we removed the variable with the highest calculated VIF value (as it is already well-explained by the remaining variables) and recalculated VIF for each remaining variable. The process was stopped when all remaining variables had a calculated $VIF \leq 5$ so that there was no concern of high collinearity among the variables.

FA Method

FA is a statistical method used to describe variability among observed and correlated variables in terms of a fewer number of unobserved and underlying variables called factors. This method aims to identify the independent factors that explain the different sources of variability in the original variables. It assumes that any observed variable is directly associated with an underlying factor. Informally, FA shares many similarities with principal component analysis: they are both data analysis techniques, their goal being to reduce a large number of variables into a fewer number of more treatable and interpretable new variables, trying to minimize the information loss at the same time. This dimensionality reduction is performed by projecting the original data onto a lower-dimensional space, preserving as much variability as possible. More formally, consider a set A of n observable variables for which we assume that each variable x_i can be expressed as a linear combination of $k < n$ factors and intercept β_i :

$$x_i = \beta_i + l_{i1}F_1 + l_{i2}F_2 + \dots + l_{ik}F_k + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$. We can rewrite the above in matrix notation:

$$X = \beta + LF + \varepsilon,$$

where $X = (x_1, \dots, x_n)^T$, $\beta = (\beta_1, \dots, \beta_n)^T$, and $L = \begin{pmatrix} l_{11} & \dots & l_{1k} \\ \vdots & & \vdots \\ l_{n1} & \dots & l_{nk} \end{pmatrix}$ denote the factor-loading matrix, $F = (F_1, \dots, F_k)^T$ is the vector of common factors, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is the vector of unobserved error terms. The factor-loading matrix L expresses the relationship of each observed variable with the unobserved factors, showing the variance explained by each observed variable for each factor. If we consider $X \in R^n$ to be the vector representing all questionnaire items, we can replace it with the vector $F \in R^k$ of unobserved factors, which lies in a lower-dimensional space. We can think of each factor as being capable of explaining a certain variance in the original items, and we can further exclude factors with the lowest amount of explained variance (the criterion for selecting the number of factors is explained in the following sections).

We looked for joint variations in the observed variables in response to unobserved latent factors. Three assumptions should be satisfied: there should be no outliers in the data, the sample size should be *big enough*, and there should be no perfect multicollinearity between the observed variables [32]. Our data set satisfied all of these assumptions.

We further applied two adequacy tests before proceeding with FA: the Bartlett test of sphericity [33] and the Kaiser–Meyer–Olkin (KMO) test [34]. The Bartlett test of sphericity checks whether the observed variables are effectively correlated with each other by comparing their correlation matrix $\text{corr}(X)$ against the identity matrix I . If the null hypothesis ($\text{corr}(X) = I$) cannot be rejected, this indicates that the original variables are orthogonal (hence, unsuitable for structure

detection), and any data reduction technique (such as FA) would not produce any meaningful result. The KMO test is another way of measuring the suitability of data for FA; it estimates the proportion of variance that may be common (ie, caused by the same underlying factor) among all observed variables, with a lower proportion being a more suitable condition for FA. KMO values range between 0 and 1 (according to the original paper by Kaiser [34], values of a statistic < 0.5 mean that performing FA is not adequate). Together, the 2 adequacy tests check the data distribution and sampling adequacy.

Finally, we used the Kaiser criterion [35] to select an adequate number of factors. The eigenvalues λ_i of the correlation matrix $\text{corr}(X)$ can be used to measure the degree to which the factors explain the variance in the observed variables. Thus, any factor with an associated eigenvalue $\lambda_i > 1$ explains more of the variance than a single (observed) variable. As a selection criterion, the number of factors was chosen such that it was equal to the number of eigenvalues > 1 . However, the Kaiser criterion has been criticized for being an arbitrary approach [36]. Thus, we further examined the scree plot of the eigenvalues and checked whether our stopping criterion matched the possible points of inflection. Finally, we used ordinary least squares to identify the minimum residual solution, which is also the default method for exploratory FA, when estimating the factor loadings.

Ethical Approval

This study was approved by the research ethics committee of the Faculty of Medicine, University of Tokyo (application number 11781).

Results

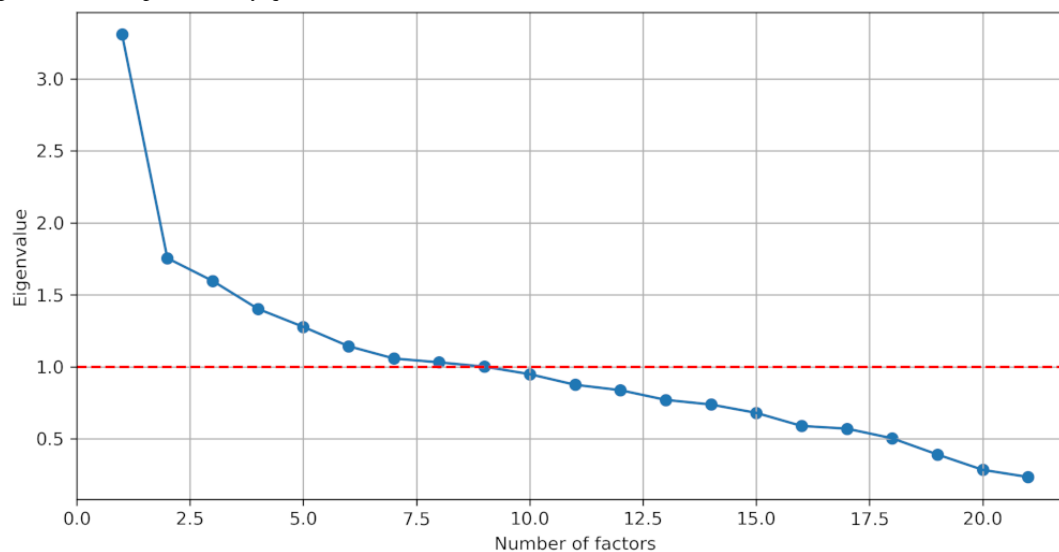
For meaningful comparisons, we grouped the questions according to the frequency at which they were asked (daily, weekly, and monthly) and analyzed each group separately.

Daily Questions

A total of 21 questions were asked daily to 79 users for a total of 1746 collected data points. After excluding missing values, of the 1746 collected data points, 1658 (95%) data points were left for analysis.

As described in the *Methods* section, we iteratively excluded all the variables for which $VIF_i > 5$. By doing so, the 21 initial variables were reduced to 3 (items 4, 20, and 21).

When FA was applied to the daily questions set, we obtained a P value of 0 using the Bartlett test; thus, we rejected the null hypothesis $\text{corr}(X) = I$ at all significance levels, confirming that the considered variables were correlated with each other. The KMO test yielded a value of 0.73, further confirming that the structure detection was adequate for the considered variables. The Kaiser criterion led to the selection of 9 factors (with the first factor having a significantly higher eigenvalue), as shown in the eigenvalue scree plot in [Figure 1](#).

Figure 1. Eigenvalues scree plot for daily questions.

By analyzing the factor-loading matrix L of the 9 factors for daily questions, we observed the following: (1) factor 1 had high factor loadings for items 2, 3, 4, and 5 (sleep-related questions); (2) factor 2 had high factor loadings for item 20 (stress level); (3) factor 3 had high factor loadings for item 6 (whether eating 3 times a day); (4) factor 4 had high factor loadings for items 16, 17, and 18 (related to caffeinated drinks); and (5) factor 5 had high factor loadings for items 9 and 12 (having lunch in a restaurant or having a takeout lunch). The other factors were less interesting (eg, as they were each associated with 1 or 2 variables only and displayed factor loadings smaller in magnitude than those of the 5 abovementioned factors) or were more difficult to interpret, as the amount of variance they explained was relatively low and spread across multiple variables. The factor-loading matrix for daily questions is provided in [Multimedia Appendix 2](#).

When the sample size is small (typically <300 , as in our analysis), it is also worth looking at the average communality of the retained items [37]. Using 9 underlying factors, we obtained an average communality of 0.492, which is an acceptable value when using Promax rotation [38], as we did for our analysis.

For the main 5 abovementioned factors, we also conducted a reliability analysis using Cronbach α to measure the internal consistency of our underrepresentation. The α coefficient for factor 1 was high (.895), indicating high internal consistency for the sleep-related items, confirming the reliability of the factor. Factors 2 and 3 had high factor loadings for a single variable; thus, we could not check any interitem internal consistency. Factors 4 and 5 displayed lower α coefficients (.470 and .536, respectively), which might be a consequence of the low number of represented original items (3 and 2, respectively) than factor 1 (4), as well as a reduced scale with respect to the sleep-related questions.

The 9 underlying factors (instead of the 21 observed variables) explained 49.25% of the total variance.

Weekly Questions

A total of 26 questions were asked weekly to 79 users. The VIF approach reduced the initial 26 questions to 21 (excluding items 28, 30, 35, 46, and 47).

When applying FA to the weekly question set, we again obtained a P value of 0 using the Bartlett test, and the KMO test produced a value of 0.69. The Kaiser criterion led to the selection of 9 factors. Similar to the eigenvalues for the daily questions, the first factor exhibited a significantly higher eigenvalue ([Figure 2](#)).

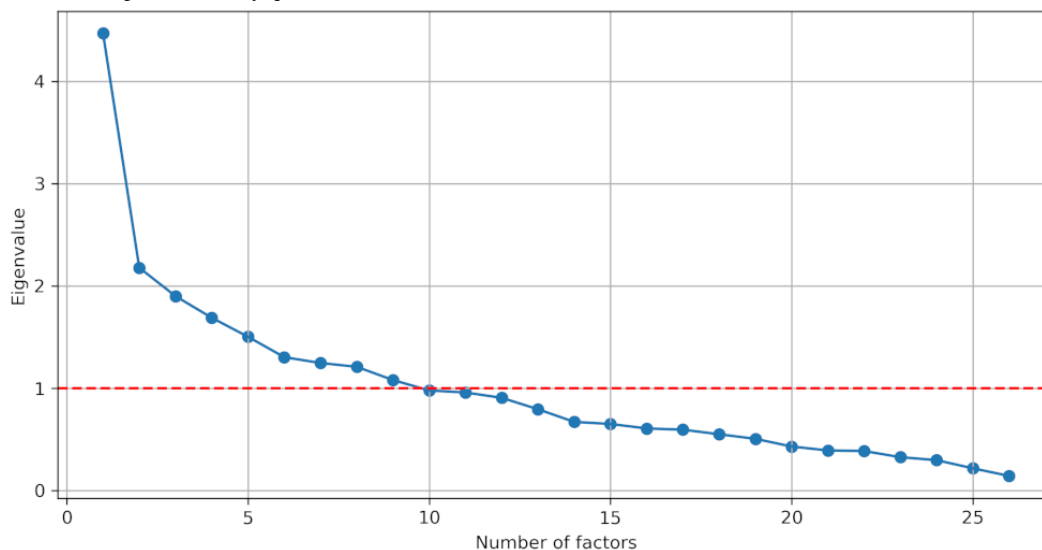
By analyzing the factor-loading matrix L of the 9 factors for weekly questions, we observed the following: (1) factor 1 had high factor loadings for items 26, 30, 31, and 39 (sleep-related questions); (2) factor 2 had high factor loadings for items 28, 34, and 35 (questions related to satisfaction and feeling well); (3) factor 3 had high factor loadings for items 32 and 33 (feeling annoyed and feeling tired during the day); (4) factor 4 had high factor loadings for items 22 and 24 (weekly number of days off and weekly hours of work); and (5) factor 5 had high factor loadings for items 23 and 46 (amount of tobacco consumption and degree of interest in a smoking cessation program). The other factors were less interesting (eg, as they were each associated with 1 or 2 variables only and displayed factor loadings smaller in magnitude than those of the 5 abovementioned factors) or were more difficult to interpret, as the amount of variance they explained was relatively low and spread across multiple variables. The factor-loading matrix for weekly questions is provided in [Multimedia Appendix 3](#).

Using 9 underlying factors, we obtained an average communality of 0.493, which is acceptable.

For the main 5 abovementioned factors, the results of the reliability analysis indicated high or acceptable Cronbach α coefficients for all considered factors (.757, .821, .721, .679, and .609), further confirming the internal consistency of the considered underrepresentation.

The 9 underlying factors (instead of the 26 observed variables) explained 49.29% of the total variance.

Figure 2. Eigenvalues scree plot for weekly questions.



Monthly Questions

A total of 4 questions were asked monthly to the 79 users. The VIF approach reduced the initial 4 questions to 2 (leaving items 50 and 51).

When applying FA to the monthly question set, we again obtained a *P* value of 0 using the Bartlett test, and the KMO test produced a value of 0.64. The Kaiser criterion led to the selection of a single factor (Figure 3).

By analyzing the factor-loading vector *L* of the factor for monthly questions, we observed that factor 1 had relatively high factor loadings for all 4 monthly questions, in particular items 50 and 51 (*degree of feeling on top of things* and *degree of not being able to cope with all the things that needed to be done*,

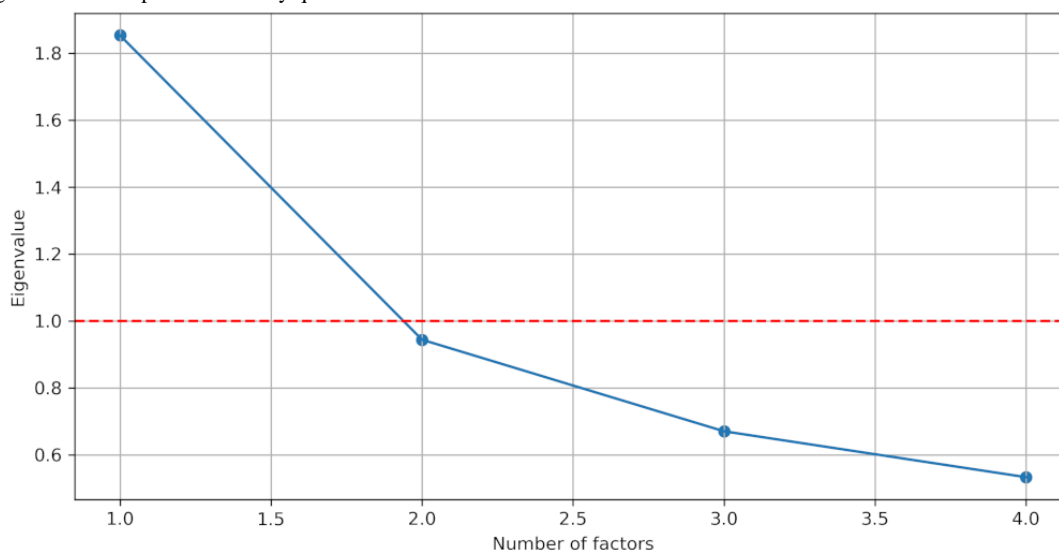
respectively). The factor-loading vector for monthly questions is provided in [Multimedia Appendix 4](#).

Using a single underlying factor, we obtained an average communality of 0.308, which is too low. However, as communalities are the proportion of each variable’s variance that can be explained by the factors (in this case, a single factor), it makes sense that their average matches the cumulative variance explained by the single factor itself.

For the single considered factor, the results of the reliability analysis indicated an acceptable Cronbach α coefficient (.607), further confirming the internal consistency of the considered underrepresentation.

The single underlying factor (instead of the 4 observed variables) explained 30.76% of the total variance.

Figure 3. Eigenvalues scree plot for monthly questions.



Discussion

Principal Findings

The objective of this study was to compare 2 different statistical methods, which are often used as reduction techniques, and

their results in a lifestyle questionnaire, which was constructed using a variety of questions (asked with daily, weekly, and monthly frequency) aimed at evaluating the general well-being of an individual. Our main findings suggest not only that existing validated lifestyle questionnaires might benefit from further

item reduction (in the questions about sleep quality and satisfaction, in particular) but also that different algorithms lead to different results for what concerns other groups of items.

Indeed, the results obtained using the two considered methods (VIF and FA) shared some similarities but also exhibited substantial differences.

Among daily questions, VIF led to the exclusion of many more variables than FA, retaining only 1 question about sleep and 2 questions aimed at evaluating stress levels at work and at home, respectively. Although we expected the sleep items to be highly correlated, the iterative approach of deleting correlated variables also led to the exclusion of all items on eating habits, suggesting that, at least in our data sample, these variables shared some correlation with sleep satisfaction and stress level components. The two types of stress do not seem to be well correlated, suggesting that stress at home does not depend on the amount of stress felt at work and vice versa. FA suggested the inclusion of only 9 factors instead of the original 21 variables; however, in line with the VIF, the factor with the highest eigenvalue captured the variance of the sleep-related items.

Among the weekly questions, the VIF reduced the initial 26 weekly questions to 21 questions. Among the 5 deleted questions, 3 (60%) were sleep-related, and it made sense that they were correlated with the other weekly sleep items. However, interestingly, there was also a correlation between desiring a healthier lifestyle (identified as an interest in quitting smoking and consideration for reducing alcohol intake) and work habits or sleep satisfaction level. This is in line with findings from Hidaka et al [39], who observed a positive correlation between low sleep satisfaction and unhealthy lifestyle patterns in the Japanese population. FA produced similar results, identifying 1 factor in particular for sleep-related questions and another factor for general well-being.

Finally, the monthly questions selected from the Perceived Stress Scale-4 aimed to assess different aspects of the consequences of stress. For example, stress is known to negatively affect self-confidence [40] and, in the long-run, also mental and physical health [41]. Therefore, we expected that the analyses would lead to the exclusion of a significant number of these items. Indeed, the VIF reduced the initial 4 questions to 2, and the FA identified a single significant factor.

We observed that despite the fact that both methods deal with the same underlying problem of multicollinearity, VIF led to greater item reduction in some instances, whereas FA did so in others. Thus, we suggest that questionnaire designers use both methods and, in the event of a discrepancy in results, adopt other additional measures such as comparing both results with the consistency of the internal questionnaire obtained using Cronbach α for the final selection.

As previously mentioned, our questionnaire included 2 reverse-coded questions. In our methods, we analyzed the

absolute value of the correlation coefficients to eliminate the need for any additional operations to identify the direction of the correlation (positive or negative).

The principal result of our study shows that even in validated lifestyle questionnaires, many items (particularly sleep-related ones) are indeed redundant. Therefore, when aiming for short questionnaires, we suggest that questionnaire designers should always consider the application of item reduction instruments after a trial phase, as certain items could, in principle, be deleted without incurring significant information loss.

Our study had some limitations that warrant mention. First, because of factor loadings, FA can also be used to analyze the amount of information that is lost when switching from the original variables to the underlying factors. In general, it should be noted that the exclusive use of statistical methods to shorten questionnaires can lead to the loss of valuable information [42]. Indeed, this was observed in our results; the total variance explained exclusively by the identified factors never exceeded 50%. Therefore, rather than providing a recipe that is indiscriminately valid, our approach was to focus on identifying areas in which there is a high possibility of reducing questionnaire items with as little information loss as possible. Our empirical results show that sleep-related questions are, by far, the area where such a reduction seems the most feasible. Second, as this was a pilot study, the number of participants was relatively small. This is the major limitation of our study, as a bigger data set would provide more statistically sound results. However, despite its small size, the size of our data set was considered acceptable for the analyses we conducted [43,44]. Furthermore, despite differences in age and sex, the participants were all selected from the same company, which could have introduced some selection bias. Thus, the results obtained in this study should be validated in further studies.

Conclusions

We constructed a lifestyle questionnaire by combining items from various authoritative sources. We then applied two different statistical methods for item reduction (VIF and FA) to check whether the existing items in the three groups of questions (asked with daily, weekly, and monthly frequency) were redundant. The results of the applied methods did not always match but nevertheless provided evidence that many items related to sleep, in particular, were indeed redundant. Two reduced questionnaires (according to VIF and FA) are proposed in [Multimedia Appendix 5](#). We also conducted reliability analyses for each group of questions using Cronbach α to measure the consistency of the obtained underrepresentation, obtaining satisfactory results. Our results suggest that questionnaire designers should always conduct a *trial* phase on a sample of participants, and examine the correlation between the items, before finalizing any lifestyle questionnaire.

Acknowledgments

This research was supported by the Center of Innovation Program of the Japan Science and Technology Agency (grant number JPMJCE1304). The funders had no role in the design of the study; collection, analysis, and interpretation of data; writing of the

report; or decision to submit the paper for publication. The authors would like to thank all staff members at the Center of Innovation, University of Tokyo, for their extensive efforts and help in conducting the study. The authors would also like to thank all members of Precision Health, The University of Tokyo, for their invaluable assistance and, in particular, Dr Masahiro Nakamura for his help with the preparation of the data set. The authors wish to thank the Nomura Securities Health Insurance Association for recruiting participants and collecting data. The authors also wish to thank Shinya Uezono (SoftBank Corp), Shunya Arita (SoftBank Corp), Toshiki Murata (SoftBank Corp), Saeru Yamamuro (SoftBank Corp), Tomoki Kinoue (SoftBank Corp), Tomihisa Ikeura (The University of Tokyo, Center of Innovation Program), and Yoshifumi Matsui (Albert Inc) for their support in this project. The authors would like to thank Tatsunori Seki (SoftBank Corp) and Yoshiya Shigeeda (SoftBank Corp) for their insightful comments and suggestions.

Authors' Contributions

TS, AKS, and AS designed the main concepts of this study. AS and KF analyzed and interpreted the data. AS drafted the paper. TS, KF, AKS, and UC critically revised the manuscript for important intellectual content. All the authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Lifestyle questionnaire.

[\[XLSX File \(Microsoft Excel File\), 13 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Factor loading matrix for daily questions.

[\[PNG File , 69 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Factor loading matrix for weekly questions.

[\[PNG File , 85 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Factor loading vector for monthly questions.

[\[PNG File , 3 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Proposed reduced lifestyle questionnaires.

[\[XLSX File \(Microsoft Excel File\), 12 KB-Multimedia Appendix 5\]](#)

References

1. Lavrakas PJ. Encyclopedia of survey research methods. Thousand Oaks, CA: Sage Publications; 2008.
2. Herzog AR, Bachman JG. Effects of questionnaire length on response quality. *Public Opin Q* 1981;45(4):549-559. [doi: [10.1086/268687](https://doi.org/10.1086/268687)]
3. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. *J Exp Soc Psychol* 2009;45(4):867-872. [doi: [10.1016/j.jesp.2009.03.009](https://doi.org/10.1016/j.jesp.2009.03.009)]
4. Maloney P, Grawitch MJ, Barber LK. Strategic item selection to reduce survey length: reduction in validity? *Consult Psychol J Pract Res* 2011;63(3):162-175. [doi: [10.1037/a0025604](https://doi.org/10.1037/a0025604)]
5. Stanton JM, Sinar EF, Balzer WK, Smith PC. Issues and strategies for reducing the length of self-report scales. *Pers Psychol* 2002;55(1):167-194. [doi: [10.1111/j.1744-6570.2002.tb00108.x](https://doi.org/10.1111/j.1744-6570.2002.tb00108.x)]
6. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16(3):297-334. [doi: [10.1007/BF02310555](https://doi.org/10.1007/BF02310555)]
7. McHorney CA, Ware Jr JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31(3):247-263. [doi: [10.1097/00005650-199303000-00006](https://doi.org/10.1097/00005650-199303000-00006)] [Medline: [8450681](https://pubmed.ncbi.nlm.nih.gov/8450681/)]

8. Bai Y, Li J, Bai Y, Ma W, Yang X, Ma F. Development and validation of a questionnaire to evaluate the factors influencing training transfer among nursing professionals. *BMC Health Serv Res* 2018;18(1):107 [FREE Full text] [doi: [10.1186/s12913-018-2910-7](https://doi.org/10.1186/s12913-018-2910-7)] [Medline: [29433487](https://pubmed.ncbi.nlm.nih.gov/29433487/)]
9. Brosnan K, Grün B, Dolnicar S. Identifying superfluous survey items. *J Retail Consum Serv* 2018;43:39-45. [doi: [10.1016/j.jretconser.2018.02.007](https://doi.org/10.1016/j.jretconser.2018.02.007)]
10. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res* 2007;16 Suppl 1:5-18. [doi: [10.1007/s11136-007-9198-0](https://doi.org/10.1007/s11136-007-9198-0)] [Medline: [17375372](https://pubmed.ncbi.nlm.nih.gov/17375372/)]
11. Ledbetter MK, Diawara N, Porter BE. The use of item response theory in survey methodology: application in seat belt data. *Am J Oper Res* 2018;8:17-32. [doi: [10.4236/ajor.2018.81002](https://doi.org/10.4236/ajor.2018.81002)]
12. Allen DS. The impact of shortening a long survey on response rate and response quality. Brigham Young University. 2016. URL: <https://scholarsarchive.byu.edu/etd/5968> [accessed 2022-03-03]
13. Liao D, Valliant R. Variance inflation factors in the analysis of complex survey data. *Surv Methodol* 2012;38(1):53-62.
14. Liao D. Collinearity diagnostics for complex survey data. University of Maryland. 2010. URL: <https://drum.lib.umd.edu/handle/1903/10881> [accessed 2022-03-03]
15. Roberts JG, Tugwell P. Comparison of questionnaires determining patient satisfaction with medical care. *Health Serv Res* 1987;22(5):637-654 [FREE Full text] [Medline: [3692863](https://pubmed.ncbi.nlm.nih.gov/3692863/)]
16. Wilson A, McDonald P. Comparison of patient questionnaire, medical record, and audio tape in assessment of health promotion in general practice consultations. *BMJ* 1994;309(6967):1483-1485 [FREE Full text] [doi: [10.1136/bmj.309.6967.1483](https://doi.org/10.1136/bmj.309.6967.1483)] [Medline: [7804055](https://pubmed.ncbi.nlm.nih.gov/7804055/)]
17. Dell-Kuster S, Sanjuan E, Todorov A, Weber H, Heberer M, Rosenthal R. Designing questionnaires: healthcare survey to compare two different response scales. *BMC Med Res Methodol* 2014;14:96 [FREE Full text] [doi: [10.1186/1471-2288-14-96](https://doi.org/10.1186/1471-2288-14-96)] [Medline: [25086869](https://pubmed.ncbi.nlm.nih.gov/25086869/)]
18. LaDonna KA, Taylor T, Lingard L. Why open-ended survey questions are unlikely to support rigorous qualitative insights. *Acad Med* 2018;93(3):347-349. [doi: [10.1097/ACM.0000000000002088](https://doi.org/10.1097/ACM.0000000000002088)] [Medline: [29215376](https://pubmed.ncbi.nlm.nih.gov/29215376/)]
19. Cappelleri JC, Gerber RA, Kourides IA, Gelfand RA. Development and factor analysis of a questionnaire to measure patient satisfaction with injected and inhaled insulin for type 1 diabetes. *Diabetes Care* 2000;23(12):1799-1803. [doi: [10.2337/diacare.23.12.1799](https://doi.org/10.2337/diacare.23.12.1799)] [Medline: [11128356](https://pubmed.ncbi.nlm.nih.gov/11128356/)]
20. Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol* 1997;50(3):233-238. [doi: [10.1016/s0895-4356\(96\)00377-0](https://doi.org/10.1016/s0895-4356(96)00377-0)] [Medline: [9120521](https://pubmed.ncbi.nlm.nih.gov/9120521/)]
21. Arifin WN, Yusoff MS. Confirmatory factor analysis of the Universiti Sains Malaysia emotional quotient inventory among medical students in Malaysia. *SAGE Open* 2016;6(2):215824401665024. [doi: [10.1177/2158244016650240](https://doi.org/10.1177/2158244016650240)]
22. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes* 2003;1:27 [FREE Full text] [doi: [10.1186/1477-7525-1-27](https://doi.org/10.1186/1477-7525-1-27)] [Medline: [12952544](https://pubmed.ncbi.nlm.nih.gov/12952544/)]
23. Erhart M, Hagquist C, Auquier P, Rajmil L, Power M, Ravens-Sieberer U, European KIDSCREEN Group. A comparison of Rasch item-fit and Cronbach's alpha item reduction analysis for the development of a quality of life scale for children and adolescents. *Child Care Health Dev* 2010;36(4):473-484. [doi: [10.1111/j.1365-2214.2009.00998.x](https://doi.org/10.1111/j.1365-2214.2009.00998.x)] [Medline: [19702637](https://pubmed.ncbi.nlm.nih.gov/19702637/)]
24. Ellis BW, Johns MW, Lancaster R, Raptopoulos P, Angelopoulos N, Priest RG. The St. Mary's hospital sleep questionnaire: a study of reliability. *Sleep* 1981;4(1):93-97. [doi: [10.1093/sleep/4.1.93](https://doi.org/10.1093/sleep/4.1.93)] [Medline: [7232974](https://pubmed.ncbi.nlm.nih.gov/7232974/)]
25. Lee KA. Self-reported sleep disturbances in employed women. *Sleep* 1992;15(6):493-498. [doi: [10.1093/sleep/15.6.493](https://doi.org/10.1093/sleep/15.6.493)] [Medline: [1475563](https://pubmed.ncbi.nlm.nih.gov/1475563/)]
26. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. *The social psychology of health: the Claremont symposium on applied social psychology*. Thousand Oaks, CA: Sage Publications; 1988:31-67.
27. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*. 2nd edition. Boston, MA: Springer; 2012.
28. Johnston R, Jones K, Manley D. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual Quant* 2018;52(4):1957-1976 [FREE Full text] [doi: [10.1007/s11135-017-0584-6](https://doi.org/10.1007/s11135-017-0584-6)] [Medline: [29937587](https://pubmed.ncbi.nlm.nih.gov/29937587/)]
29. De Jongh PJ, De Jongh E, Pienaar M, Gordon-Grant H, Oberholzer M, Santana L. The impact of pre-selected variance inflation factor thresholds on the stability and predictive power of logistic regression models in credit scoring. *Orion* 2015;31(1):17-37. [doi: [10.5784/31-1-162](https://doi.org/10.5784/31-1-162)]
30. Sheather S. *A modern approach to regression with R*. New York, NY: Springer; 2009.
31. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning: with applications in R*. 2nd edition. New York, NY: Springer; 2013.
32. Fabrigar LR, Wegener DT. *Exploratory factor analysis: understanding statistics*. Oxford, UK: Oxford University Press; 2011.
33. Bartlett MS. Tests of significance in factor analysis. *Br J Stat Psychol* 1950;3(2):77-85. [doi: [10.1111/j.2044-8317.1950.tb00285.x](https://doi.org/10.1111/j.2044-8317.1950.tb00285.x)]
34. Kaiser HF. An index of factorial simplicity. *Psychometrika* 1974;39(1):31-36. [doi: [10.1007/BF02291575](https://doi.org/10.1007/BF02291575)]

35. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958;23(3):187-200. [doi: [10.1007/BF02289233](https://doi.org/10.1007/BF02289233)]
36. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods* 1999;4(3):272-299. [doi: [10.1037/1082-989X.4.3.272](https://doi.org/10.1037/1082-989X.4.3.272)]
37. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods* 1999;4(1):84-99. [doi: [10.1037/1082-989X.4.1.84](https://doi.org/10.1037/1082-989X.4.1.84)]
38. Osborne JW, Costello AB, Kellow JT. Best practices in exploratory factor analysis. In: Osborne JW, editor. *Best practices in quantitative methods*. Thousand Oaks, CA: Sage Publications; 2008:205-213.
39. Hidaka T, Kasuga H, Endo S, Masuishi Y, Kakamu T, Takeda A, et al. Are lifestyle pattern changes associated to poor subjective sleep quality?: a cross-sectional study by gender among the general Japanese population underwent specified medical check-ups in 2014 and 2015. *BMJ Open* 2020;10(12):e037613 [FREE Full text] [doi: [10.1136/bmjopen-2020-037613](https://doi.org/10.1136/bmjopen-2020-037613)] [Medline: [33328256](https://pubmed.ncbi.nlm.nih.gov/33328256/)]
40. Goette L, Bendahan S, Thoresen J, Hollis F, Sandi C. Stress pulls us apart: anxiety leads to differences in competitive confidence under stress. *Psychoneuroendocrinology* 2015;54:115-123 [FREE Full text] [doi: [10.1016/j.psyneuen.2015.01.019](https://doi.org/10.1016/j.psyneuen.2015.01.019)] [Medline: [25705012](https://pubmed.ncbi.nlm.nih.gov/25705012/)]
41. Toussaint L, Shields GS, Dorn G, Slavich GM. Effects of lifetime stress exposure on mental and physical health in young adulthood: how stress degrades and forgiveness protects health. *J Health Psychol* 2016;21(6):1004-1014 [FREE Full text] [doi: [10.1177/1359105314544132](https://doi.org/10.1177/1359105314544132)] [Medline: [25139892](https://pubmed.ncbi.nlm.nih.gov/25139892/)]
42. Jensen J, Burlingame G. An item reduction analysis of the group questionnaire. *Psychotherapy (Chic)* 2018;55(2):144-150. [doi: [10.1037/pst0000145](https://doi.org/10.1037/pst0000145)] [Medline: [29863394](https://pubmed.ncbi.nlm.nih.gov/29863394/)]
43. de Winter JC, Dodou D, Wieringa PA. Exploratory factor analysis with small sample sizes. *Multivariate Behav Res* 2009;44(2):147-181. [doi: [10.1080/00273170902794206](https://doi.org/10.1080/00273170902794206)] [Medline: [26754265](https://pubmed.ncbi.nlm.nih.gov/26754265/)]
44. Jung S, Lee S. Exploratory factor analysis for small samples. *Behav Res Methods* 2011;43(3):701-709. [doi: [10.3758/s13428-011-0077-9](https://doi.org/10.3758/s13428-011-0077-9)] [Medline: [21431996](https://pubmed.ncbi.nlm.nih.gov/21431996/)]

Abbreviations

- FA:** factor analysis
- IRT:** Item Response Theory
- KMO:** Kaiser–Meyer–Olkin
- VIF:** variance inflation factor

Edited by T Leung; submitted 05.04.21; peer-reviewed by A Chang, R Watson, M Nieroda; comments to author 26.04.21; revised version received 18.06.21; accepted 09.02.22; published 18.03.22

Please cite as:

Staffini A, Fujita K, Svensson AK, Chung UI, Svensson T
Statistical Methods for Item Reduction in a Representative Lifestyle Questionnaire: Pilot Questionnaire Study
Interact J Med Res 2022;11(1):e28692
URL: <https://www.i-jmr.org/2022/1/e28692>
doi: [10.2196/28692](https://doi.org/10.2196/28692)
PMID:

©Alessio Staffini, Kento Fujita, Akiko Kishi Svensson, Ung-Il Chung, Thomas Svensson. Originally published in the Interactive Journal of Medical Research (<https://www.i-jmr.org/>), 18.03.2022. 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.