



OPEN ACCESS

EDITED BY

César González-Blanch,
Marqués de Valdecilla University Hospital, Spain

REVIEWED BY

Cristian Ramos-Vera,
Cesar Vallejo University, Peru
Itumeleng P. Khumalo,
University of Johannesburg, South Africa

*CORRESPONDENCE

Eun-Hyun Lee
✉ ehlee@ajou.ac.kr

RECEIVED 11 May 2023

ACCEPTED 16 August 2023

PUBLISHED 31 August 2023

CITATION

Lee E-H, Kang EH, Kang H-J and Lee HY (2023) Measurement invariance of the patient health questionnaire-9 depression scale in a nationally representative population-based sample. *Front. Psychol.* 14:1217038. doi: 10.3389/fpsyg.2023.1217038

COPYRIGHT

© 2023 Lee, Kang, Kang and Lee. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Measurement invariance of the patient health questionnaire-9 depression scale in a nationally representative population-based sample

Eun-Hyun Lee^{1*}, Eun Hee Kang¹, Hyun-Jung Kang¹ and Hyun Young Lee²

¹Graduate School of Public Health, Ajou University, Suwon, Republic of Korea, ²Clinical Trial Center, Ajou University Hospital, Suwon, Republic of Korea

The Patient Health Questionnaire-9 (PHQ-9) is widely used to measure the severity of depressive symptoms and to screen for depressive disorder, but its measurement invariance has received little research attention. The aim of this study was to assess the measurement invariance of the PHQ-9 across various sociodemographic and medical-condition groups. The structural validity and internal consistency of the PHQ-9 were also assessed as the prerequisite properties for measurement invariance. This study was conducted using data from the Korea National Health and Nutrition Examination Survey. The included participants comprised 5,347 people older than 19 years. Exploratory graph analysis (EGA) and confirmatory factor analysis (CFA) were performed to determine structural validity, and the omega coefficient (ω) was used to determine internal consistency. Measurement invariance (configural, metric, and scalar invariance) was evaluated using multigroup confirmatory factor analysis (MGCFA). The single structural model of the PHQ-9 that was validated by EGA was also satisfactory with fit indices of $\chi^2 = 770.765$ ($p < 0.001$), CFI = 0.944, SRMR = 0.040, and RMSEA = 0.076 (90% CI = 0.072–0.081). The ω of the PHQ-9 was 0.812, implying satisfactory internal consistency. The one-factor PHQ-9 had equivalent overall structure, factor loadings, and item intercepts across age groups, suggesting invariance across ages. Partial scalar invariance was demonstrated across sex and marital-status groups. Partial metric and scalar invariance were supported across education groups. Scalar invariance was supported among all of the medical-condition (hypertension, diabetes, cancer, arthritis, asthma, and heart disease) groups. Overall, the measurement invariance of the one-factor PHQ-9 was empirically supported across sociodemographic and medical-condition groups. The PHQ-9 can be reliably used to compare the severity of depressive symptoms across these groups in research and practice.

KEYWORDS

depressive symptoms, measurement invariance, network psychometrics, structural validity, internal consistency, questionnaire

Introduction

Depression is a common public health concern. It has been estimated that about 5% of adults suffer from depression worldwide, which impairs their daily functioning at work or in the family and adversely affects the quality of life, and may even result in suicide (World Health Organization, 2021). It also brings large economic costs related directly to the workplace (absenteeism and presenteeism) and to suicide (Greenberg et al., 2021). Early detection and prompt treatment are therefore important. For these reasons, the United States Preventive Services Task Force recommended screening for depression in the general adult population (Siu et al., 2016).

The Patient Health Questionnaire-9 (PHQ-9) is a self-administered instrument that was developed to identify people who may have depression and assess the severity of depression symptoms in research and primary care (Kroenke et al., 2001). The PHQ-9 comprises nine items corresponding to the nine diagnostic criteria for depressive disorder in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association, 2000). When the PHQ-9 was initially developed using 6,000 patients across 8 primary-care and 7 obstetrics-gynecology clinics, it demonstrated satisfactory internal consistency, test-retest reliability, and convergent validity. Using a score criterion of ≥ 10 for major depression produced a sensitivity of 88% and specificity of 88%. The PHQ-9 was subsequently psychometrically examined in diverse populations (e.g., people with chronic disease, the elderly, college students, and adults) and languages (El-Den et al., 2018; Carroll et al., 2020). The PHQ-9 has been considered one of the most widely used self-reported measures in various primary-care settings worldwide (El-Den et al., 2018). The PHQ-9 has also been used in nationally representative health surveys, such as the National Health and Nutrition Examination Survey (NHANES) in the United States (Centers for Disease Control and Prevention, 2020), the Peruvian Demographic and Health Survey (Villarreal-Zegarra et al., 2019), the UK Biobank (Davis et al., 2020), and the Korea National Health and Nutrition Examination Survey (KNHANES) in South Korea (Korea Centers for Disease Control and Prevention, 2022).

Despite the widespread use of the PHQ-9, its psychometric property of measurement invariance has received little attention (Teymoori et al., 2020). Measurement invariance refers to the psychometric equivalence of a construct across groups and demonstrates that the construct has the same meaning among groups (Polit and Yang, 2016). Evidence for measurement invariance across the groups is needed when comparing differences in PHQ-9 scores between groups in research and practice, since a lack of such evidence could mean that the group difference is due to other measurement aspects rather than representing a true difference in depressive symptoms (Patel et al., 2019). This may result in the under- or over-detection of people with depression in certain groups.

The measurement invariance of the PHQ-9 has been evaluated across sociodemographic groups (e.g., sex, age, marital status, education level, employment status, and race/ethnicity groups) in the United States, Spain, Germany, Bangladesh, and Portugal (Petersen et al., 2015; González-Blanch et al., 2018; Patel et al., 2019; Villarreal-Zegarra et al., 2019; Lamela et al., 2020; Rahman et al., 2022). The measurement invariance of the PHQ-9 has rarely been evaluated across medical-condition groups, so it remains unclear whether the meanings of the PHQ-9 items are similar between groups with and

without a specific medical condition (e.g., disease). Nevertheless, differences in the PHQ-9 among medical-condition groups have been continuously reported. For example, depressive symptoms measured using the PHQ-9 were compared between patients with cancer and a general population (Hinz et al., 2016), between normotensive and prehypertensive groups (Jang, 2021), and between nonstroke and stroke survivor groups (Hong et al., 2021) without the evidence of measurement invariance in the PHQ-9 across these groups. In other words, the findings of group differences in depressive symptoms were insufficient for the confidence that the differences were due to the true nature of depression. Two recent studies evaluated the measurement invariance of the PHQ-9 across people with and without diabetes (Nouwen et al., 2021) and adults with and without HIV (Mwangi et al., 2020). More validation is needed to determine the measurement invariance of the PHQ-9 across groups with various medical conditions.

The aim of this study was therefore to determine the measurement invariance of the PHQ-9 across various sociodemographic (age, sex, marital status, and education level) and medical condition (hypertension, diabetes, cancer, arthritis, asthma, and heart disease) groups using a nationally representative Korean database. According to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN), evidence of structural validity and internal consistency in a self-reported instrument are prerequisites when examining measurement invariance (Prinsen et al., 2018). Thus, the structural validity and internal consistency of the PHQ-9 were also assessed as the prerequisite properties for measurement invariance.

Methods

Study design and participants

A secondary data analysis study was conducted to psychometrically evaluate the PHQ-9 using a dataset from the KNHANES VIII-2, which was a cross-sectional nationwide survey conducted by the Korea Centers for Disease Control and Prevention (KCDC) (Korea Centers for Disease Control and Prevention, 2022). The KNHANES VIII-2 used a stratified multistage cluster sampling method to obtain a sample representative of the population. The protocol used for the survey comprised three components: (a) physical and laboratory examinations by a health professional in a mobile examination car, (b) a health survey administered using interviews and self-reported questionnaires in a mobile examination car, and (c) a nutritional survey using interviews in home visits. The first-two components were used in the present study. The KNHANES VIII-2 investigated 7,359 people from 3,314 households in 180 survey districts, 7,096 of whom participated in the first two components (Korea Centers for Disease Control and Prevention, 2022). Of them, 5,347 people who were older than 19 years and had completed at least 80% of the PHQ-9 items were finally included in the present study.

Ethical considerations

Data collection for the KNHANES VIII-2 was permitted by an institutional review board of the KCDC (approval no. 2018-01-03-2C-A). The data were publicly released in 2022. This study was exempted

from requiring informed consents by the Institutional Review Board at Ajou University Hospital (approval no. AJOURB-EX-2022-397).

Measures

PHQ-9

The PHQ-9 (Kroenke et al., 2001) is a self-administered instrument that was developed to screen or assess the severity of depressive symptoms in primary-care settings. Each item is scored on a 4-point Likert scale with response options from 0 (“not at all”) to 3 (“nearly every day”) that refer to events during the previous 2 weeks. Total scores range from 0 to 27, with higher scores indicating more-severe depressive symptoms. The PHQ-9 demonstrated satisfactory internal consistency (Cronbach’s $\alpha = 0.89$) and good sensitivity and specificity in identifying cases of major depressive disorders. The PHQ-9 was administered to obtain data for the KNHANES VIII-2 using face-to-face interviews. Sample weighting was not applied in the present study because the aim was to determine the measurement invariance of the PHQ-9.

Sociodemographic variables

Data on age, sex, marital status, and education level were collected from the self-reported health survey data set of the KNHANES VIII-2. Age was grouped into <65 and ≥ 65 years; sex into male and female; marital status into living with a spouse, divorced/widowed/separated, and never married/single; and education level into graduated from elementary school, middle school, high school, and college or above.

Medical condition variables

In the physical laboratory examinations of the KNHANES VIII-2, hypertension was defined as a systolic blood pressure of ≥ 140 mmHg, a diastolic blood pressure of ≥ 90 mmHg, taking a medication for high blood pressure. Prehypertension was defined as a systolic blood pressure of ≥ 120 and < 140 mmHg or a diastolic blood pressure of ≥ 80 and < 90 mmHg. Normal was defined as a systolic blood pressure of < 120 mmHg and a diastolic blood pressure of < 80 mmHg. Diabetes was diagnosed as a fasting blood glucose level of ≥ 126 mg/L, receiving a hypoglycemic agent/insulin injection, being diagnosed by a doctor, or having HbA1c $\geq 6.5\%$. Prediabetes was defined as a fasting blood glucose level of ≥ 100 and ≤ 125 mg/L or HbA1c ≥ 5.7 and $< 6.4\%$. Normal was defined as a fasting blood glucose level of < 100 mg/L or HbA1c $< 5.7\%$. Other medical-condition variables (cancer, arthritis, asthma, and heart disease) were classified into groups with and without the disease, based on the self-reported physician diagnosis in the health survey data set of the KNHANES VIII-2.

Statistical analysis

Data were analyzed using SPSS for Windows (version 25), AMOS software (version 25), and the *EGAnet* package in the R environment. For the cross-validation approach to the structural validity of the PHQ-9, the total data were split into two subsamples using the random assignment function of SPSS. Subsample 1 ($n = 2,673$) was used for exploratory graph analysis (EGA) using the *EGAnet* package to investigate the underlying dimensionality of the PHQ-9. EGA is a recently developed method for investigating the number of

dimensions (Golino and Epskamp, 2017). EGA was applied in the present study using the graphical least absolute shrinkage and selection operator (LASSO) method with the Walktrap community detection algorithm. This process graphed a network model, and edge weights were calculated. The graphical model was visually presented using nodes (items) and edges (links) that indicated correlations between two nodes after controlling for all other nodes in the network. EGA detects the number of dimensions (communities) by arranging densely clustered nodes into each dimension. The nodes are colored according to their identified communities. The network model is visually depicted with the weight matrices represented by the edges between the nodes.

The detected dimensionality of the PHQ-9 was further assessed through a nonparametric bootstrap procedure with 1,000 iterations using the *bootEGA* function. This analysis calculated the structural consistency (the proportion of times that each dimension derived from EGA is exactly recovered from the replicate bootstrap samples) and item stability (the proportion of times that a given item belongs to the same dimension obtained in the EGA in the bootstrap replications) (Golino and Christensen, 2022). Network loadings, which refers the association of each node with the dimension in which it belongs, were then obtained by using the *net.loads* function and interpreted as small (0.00–0.15), moderate (0.16–0.25), or large (0.26–0.35) loadings (Christensen and Golino, 2021a).

Confirmatory factor analysis (CFA) was subsequently conducted using AMOS to test the fit of the structure identified in the PHQ-9 network by the EGA in subsample 2 ($n = 2,674$). The CFA model was estimated using 1,000 bootstrap samples due to the assumption of the multivariate normality not being satisfied based on a Mardia’s coefficient estimate of > 5.00 (Byrne, 2016). The goodness of fit of the model was assessed using multiple indices: comparative fit index (CFI), standardized root-mean-square residual (SRMR), and root-mean-square error of approximation (RMSEA). CFI values greater than 0.95 indicate a good fit, and values of 0.90–0.95 are considered to indicate an acceptable fit (Hu and Bentler, 1999). RMSEA and SRMR values less than 0.05 indicate a good fit, and those of 0.05–0.08 indicate an adequate fit (MacCallum et al., 1996). The traditional χ^2 value and the number of degrees of freedom were also reported, but they were not used to determine the model fit since they are sensitive to the sample size (Schreiber, 2008).

With the total sample, the internal consistency of the PHQ-9 was assessed using omega coefficient (ϵ) with a criterion value of > 0.70 (McDonald, 1999). Measurement invariance of the PHQ-9 across various sociodemographic and medical-condition groups was tested using multigroup CFA (MG-CFA). There are four levels of invariance tests that progress in a hierarchical bottom-up approach: (a) configural invariance, when the number of latent constructs and the specific items loaded on them are assumed to be equivalent across groups, (b) metric invariance, when factor loadings from items to factors are assumed to be equal across groups, (c) scalar invariance, when factor loadings and item intercepts are assumed to be equal across groups, and (d) error variance invariance, when the error terms of items are assumed to be equal across groups in addition to the equality of the scalar invariance. The error variance invariance is considered to be excessively stringent and is often not achieved in practice (Chen and Tang, 2006), and so the first three levels of invariance tests were successively conducted in this study. In each progression, the higher model level was accepted if the value from the CFA decreased by

<0.010, supplemented by changes in RMSEA (Δ RMSEA) of <0.015 and SRMR (Δ SRMR) of <0.030 (for metric invariance) or <0.150 (for scalar invariance) (Chen, 2007). If a full metric or scalar invariance was not met, partial invariance was tested using the process of freeing factor loadings or item intercepts to detect noninvariant items.

31.9% ($n = 1,705$) and 15.1% ($n = 809$), respectively. Participants who reported that they had been diagnosed with cancer, arthritis, asthma, and heart disease by a physician comprised 5.7, 12.9, 3.4, and 3.2%, respectively.

Results

Descriptive statistics of study variables

The characteristics of the 5,347 participants are listed in Table 1. They comprised 54.8% ($n = 2,931$) females, and 15.2% ($n = 1,349$) were aged 65 years and older (age for the total sample = 51.26 ± 17.04 years). Participants living with their spouse comprised 65.2% ($n = 3,486$), and those with college education or above comprised 28.3% ($n = 2,050$). Participants diagnosed with hypertension and diabetes comprised

Prerequisites for measurement invariance: structural validity

Dimensionality by EGA with subsample 1

The EGA detected one dimension (communality) of nodes that are depicted using identical colors in Figure 1, suggesting that one dimension contained all nine items. The edge weights (partial correlations between nodes) are presented in Supplementary Table S1. The highest edge weight was between items 6 and 9; in other words, these item pairs exhibited relatively stronger associations. The dimensionality structure of the PHQ-9 using *bootEGA* indicated that

TABLE 1 Characteristics of the study participants.

Variable	Groups	Total sample ($n = 5,347$)	Subsample 1 ($n = 2,673$)	Subsample 2 ($n = 2,674$)
		n (%)	n (%)	n (%)
Sex	Male	2,416 (45.2)	1,198 (44.8)	1,218 (45.4)
	Female	2,931 (54.8)	1,475 (55.2)	1,456 (54.5)
Age	<65 years	3,998 (74.8)	1,989 (74.4)	2,009 (75.1)
	≥ 65 years	1,349 (25.2)	684 (25.6)	665 (24.9)
Marital status	Living with a spouse	3,486 (65.2)	1,721 (64.4)	1,765 (66.0)
	Divorced/widowed/separated	741 (13.9)	385 (14.4)	356 (13.3)
	Never married/single	1,117 (20.9)	567 (21.2)	550 (20.6)
Education level	Graduated from elementary school	899 (16.8)	448 (16.8)	451 (16.9)
	Graduated from middle school	517 (9.7)	249 (9.3)	268 (10.0)
	Graduated from high school	1,880 (35.2)	940 (35.2)	940 (35.2)
	Graduated from college or above	2,050 (28.3)	1,036 (38.8)	1,014 (37.9)
Hypertension	Normal	2,177 (40.7)	1,087 (40.7)	1,090 (40.8)
	Prehypertension	1,406 (26.3)	700 (26.2)	706 (26.4)
	Hypertension	1,705 (31.9)	855 (32.0)	850 (31.8)
Diabetes	Normal	2,221 (41.5)	1,100 (41.2)	1,121 (41.9)
	Prediabetes	2,101 (39.3)	1,055 (39.5)	1,046 (39.1)
	Diabetes	809 (15.1)	408 (15.3)	401 (15.0)
Cancer ^a	No	5,043 (94.3)	2,509 (93.9)	2,534 (94.8)
	Yes	304 (5.7)	164 (6.1)	140 (5.2)
Arthritis	No	4,659 (87.1)	2,322 (86.9)	2,337 (87.4)
	Yes	688 (12.9)	351 (13.1)	337 (12.6)
Asthma	No	5,164 (98.0)	2,585 (96.7)	2,579 (96.4)
	Yes	183 (3.4)	88 (3.3)	95 (3.6)
Heart disease ^b	No	5,175 (96.8)	2,587 (96.8)	2,588 (96.8)
	Yes	172 (3.2)	86 (3.2)	86 (3.2)

^aStomach, liver, colon, breast, cervix, lung, thyroid, and other cancers.

^bMyocardial infarction and angina.

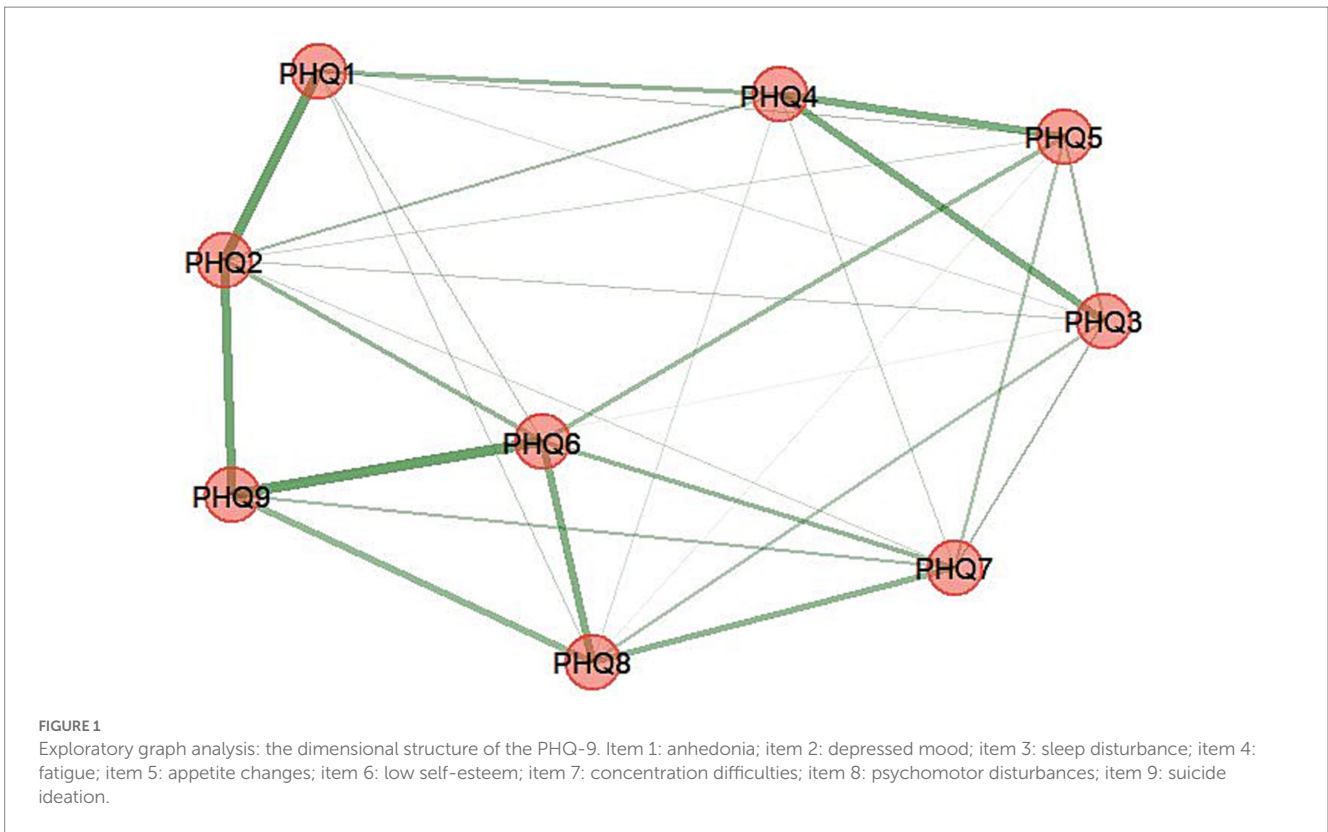


TABLE 2 Summary of fit indices in CFA of the PHQ-9.

Model	χ^2	df	CFI	SRMR	RMSEA (90% CI)	Δ CFI
Initial model	897.529*	27	0.878	0.059	0.110 (0.104–0.116)	
Modified model 1 ^a	606.745*	26	0.919	0.049	0.091 (0.085–0.098)	0.041
Modified model 2 ^b	479.763*	25	0.936	0.045	0.082 (0.076–0.089)	0.017
Modified model 3 ^c	398.205*	24	0.948	0.040	0.076 (0.070–0.083)	0.012

df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root-mean-square residual; RMSEA, root-mean-square error of approximation; Δ CFI, change in CFI compared with the previous model; CI, confidence interval.

^aCovariance between measurement errors of items 3 and 4.

^bCovariance between measurement errors of items 3 and 4, and items 7 and 8.

^cCovariance between measurement errors of items 3 and 4, items 7 and 8, and items 4 and 5.

* $p < 0.01$.

one dimension (median network structure = 1) was identified in 100% of the bootstrap iterations. That is, the structural consistency of the one-dimensional solution for the PHQ-9 was stable because the replication of >75% or more bootstrap samples is considered to exhibit adequate structural consistency (Golino et al., 2021).

If an item stability value is less than 0.80 (80%), it may be problematic (Christensen and Golino, 2021b). All of the item stability values in this study were >0.80 in the bootstrap replications (Supplementary Figure S1), and no unstable item needed to be removed from the PHQ-9. The network loading values were between 0.398 and 0.256 (Supplementary Table S2), which were interpreted as large loadings (>0.025) on their dimension.

Dimensionality by CFA with subsample 2

CFA with subsample 2 was conducted to verify the one-dimension structure of the PHQ-9 identified by EGA with subsample 1. The

initial CFA model of the PHQ-9 partially met the fit indices: $\chi^2 = 897.529$ ($p < 0.001$), CFI = 0.878, SRMR = 0.059, and RMSEA = 0.110 (90% confidence interval [CI] = 0.140–0.116). The model was therefore modified, and the model fit was substantially improved (Δ CFI > 0.10) (Byrne, 2016) when the three pairs of error terms were allowed to be correlated: $\chi^2 = 398.205$ ($p < 0.001$), CFI = 0.948, SRMR = 0.040, and RMSEA = 0.076 (90% CI = 0.070–0.083) (Table 2). Factor loading values are presented in Figure 2.

Prerequisites for measurement invariance: internal consistency

The omega coefficients of the PHQ-9 with subsample 1, subsample 2, and the total sample were 0.803, 0.821, and 0.812, respectively, implying satisfactory internal consistency.

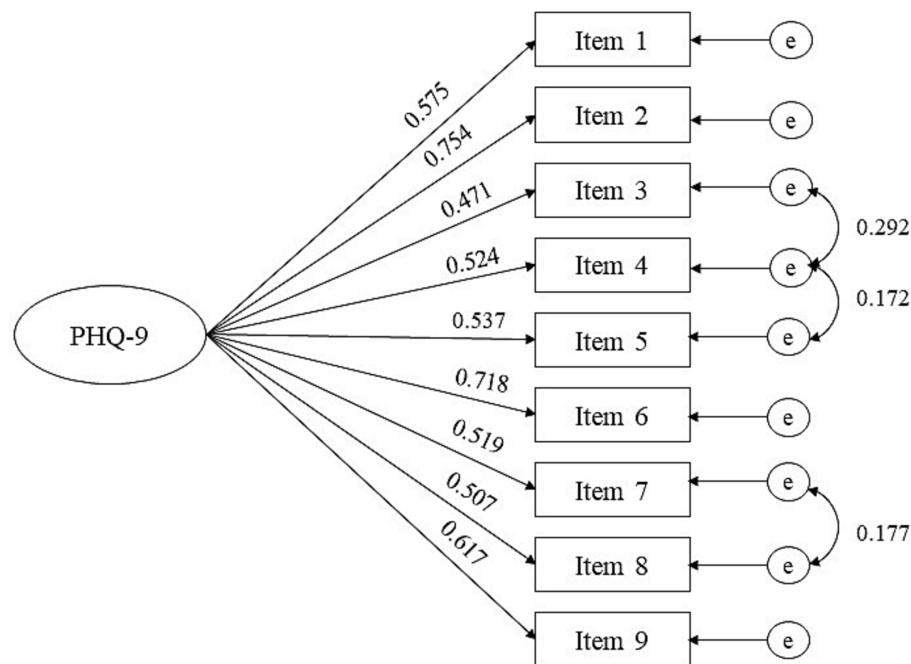


FIGURE 2

Confirmatory factor analysis model of the PHQ-9 with subsample 2. Item 1: anhedonia; item 2: depressed mood; item 3: sleep disturbance; item 4: fatigue; item 5: appetite changes; item 6: low self-esteem; item 7: concentration difficulties; item 8: psychomotor disturbances; item 9: suicide ideation; e, measurement error.

Measurement invariance

The single structural model of the PHQ-9 that was validated by EGA and CFA was also satisfactory in a total sample, with fit indices of $\chi^2 = 770.765$ ($p < 0.001$), CFI = 0.944, SRMR = 0.040, and RMSEA = 0.076 (90% CI = 0.072–0.081). The measurement invariance of the structural PHQ-9 model across each sociodemographic and medical-condition group is presented in Table 3. Regarding age, configural invariance was supported by the model fit indices (CFI = 0.939, RMSEA = 0.058, and SRMR = 0.046) across age groups (<65 vs. ≥ 65 years). Metric and scalar invariance also met the criteria for Δ CFI, Δ RMSEA, and Δ SRMR. These results demonstrated that the PHQ-9 had consistent overall structure, factor loadings, and item intercepts across age groups. Associated with sex, scalar invariance was not supported (Δ CFI = 0.012). To detect invariant item intercepts, partial scalar invariance was assessed. As a result, the partial scalar invariance model with the freely estimated item-3 intercept was supported across sex groups. There was also scalar noninvariance (Δ CFI = 0.016) in the PHQ-9 across three marital-status groups (living with a spouse vs. divorced/widowed/separated vs. never married/single). By freeing the intercepts of items 3 and 1, the partial scalar invariance model was supported across the marital groups. The metric invariance of the PHQ-9 was not supported in education level (Δ CFI = 0.014). The factor loadings of item 9 were not equal across the education-level groups. Partial metric invariance was supported by freeing the factor loadings of item 9. If partial metric invariance is not achieved, a serious measurement problem occurs and the next level of the invariance test cannot proceed (Collier, 2020). Since the partial metric invariance was achieved across education-level groups in the present study, the next level of the scalar invariance test was

performed and supported. Regarding the medical conditions (hypertension, diabetes, cancer, arthritis, asthma, and heart disease), it was found that the configural, metric, and scalar invariance of the PHQ-9 were all supported for the groups with and without each medical condition (all Δ CFI < 0.010). As an ancillary analysis, the measurement invariance of the PHQ-9 was tested across two groups: one group without disease ($n = 2,675$) and one group with at least one of hypertension, diabetes, cancer, arthritis, asthma, or heart disease ($n = 2,519$). The results supported the presence of configural, metric, and scalar invariance (all Δ CFI < 0.010).

Discussion

Prerequisites for measurement invariance: structural validity and internal consistency

When developing a self-reported scale, the most basic step is the conceptualization of the construct to be measured, and the underlying structure of that construct is assessed based on the defined conceptualization (Polit and Yang, 2016). The PHQ-9 was originally developed by turning the diagnosis criteria of the DSM-IV into self-reported items without not only conceptualization but also structural validity, even though its internal consistency, test–retest reliability, and convergent validity were satisfactory (Kroenke et al., 2001). This revealed that the factorial structure of the PHQ-9 was inconsistent. According to a systematic review, 19 of 33 studies (57.6%) examined the structural validity using a CFA-supported 1-factor structure, and 12 studies (36.4%) supported a 2-factor structure (with different types of item clustering) (Lamela et al., 2020). According to psychometric

TABLE 3 Measurement invariance of the PHQ-9 across sociodemographic and medical-condition groups.

Invariance model	χ^2	df	CFI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Age								
Configural	899.084*	48	0.939	0.058	0.046			
Metric	958.360*	56	0.934	0.055	0.047	0.005	0.004	0.001
Scalar	1062.353*	65	0.927	0.054	0.047	0.007	0.001	0.001
Sex								
Configural	863.040*	48	0.940	0.056	0.042			
Metric	927.944*	56	0.935	0.054	0.045	0.005	0.002	0.003
Scalar	1103.657*	65	0.923	0.055	0.045	0.012	0.001	<0.001
Partial scalar ^a	1045.595	64	0.927	0.054	0.045	0.008	<0.001	<0.001
Marital status								
Configural	851.518*	72	0.938	0.045	0.039			
Metric	944.781*	88	0.932	0.043	0.043	0.006	0.002	0.004
Scalar	1158.035*	106	0.916	0.043	0.043	0.016	<0.001	<0.001
Partial scalar ^b	1072.313	102	0.923	0.042	0.043	0.009	0.001	<0.001
Education level								
Configural	1114.890*	96	0.926	0.045	0.036			
Metric	1332.682*	120	0.912	0.043	0.060	0.014	0.002	0.024
Partial metric ^c	1250.648*	117	0.918	0.045	0.051	0.008	<0.001	0.015
Partial scalar ^c	1401.356*	144	0.909	0.051	0.061	0.009	0.006	0.010
Hypertension								
Configural	880.393*	72	0.940	0.046	0.046			
Metric	1031.226*	88	0.930	0.045	0.056	0.010	0.001	0.010
Scalar	1158.268*	106	0.922	0.043	0.056	0.008	0.002	<0.001
Diabetes								
Configural	896.021*	72	0.935	0.047	0.045			
Metric	958.169*	88	0.932	0.044	0.049	0.003	0.003	0.004
Scalar	1008.220*	106	0.929	0.041	0.049	0.003	0.003	<0.001
Cancer								
Configural	919.936*	48	0.936	0.058	0.041			
Metric	954.613*	56	0.934	0.055	0.041	0.002	0.003	<0.001
Scalar	967.608*	65	0.933	0.051	0.041	0.001	0.004	<0.001
Arthritis								
Configural	813.453*	48	0.943	0.055	0.039			
Metric	899.793*	56	0.937	0.053	0.040	0.006	0.001	0.001
Scalar	969.593*	65	0.932	0.051	0.041	0.003	<0.001	0.001
Asthma								
Configural	838.504*	48	0.940	0.056	0.041			
Metric	865.377*	56	0.938	0.052	0.041	0.002	0.004	0.002
Scalar	896.186*	65	0.937	0.049	0.041	0.004	0.001	<0.001
Heart disease								
Configural	849.517*	48	0.940	0.056	0.039			
Metric	915.561*	56	0.936	0.054	0.040	0.004	0.002	0.001
Scalar	930.792*	65	0.936	0.050	0.040	<0.001	0.004	<0.001

(Continued)

TABLE 3 (Continued)

Invariance model	χ^2	df	CFI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Nonexistence and existence of disease^d								
Configural	795.436*	48	0.942	0.055	0.045			
Metric	887.944*	56	0.935	0.053	0.052	0.007	0.002	0.007
Scalar	985.494*	65	0.929	0.052	0.052	0.006	0.001	<0.001

df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; Δ , change in the value compared with that in the previous model.

* $p < 0.001$.

^aItem intercept for item 3 was not constrained.

^bItem intercepts for items 3 and 1 were not constrained.

^cItem loadings for item 9 were not constrained.

^dAt least one of hypertension, diabetes, cancer, arthritis, asthma, or heart disease.

studies (Arrieta et al., 2017; Alpizar et al., 2018; Keum et al., 2018; Boothroyd et al., 2019), both one- and two-dimension models provided good fits in CFA. However, a one-dimension structure was finally chosen as providing appropriate structural validity of the PHQ-9 because the correlation between factors in the two-dimension structure was high ($\rho > 0.80$). Stochl et al. (2022) stated that the structural inconsistency is mainly due to different sample properties and methodologies. The few studies that have examined the structural validity of the PHQ-9 in South Korean populations also demonstrated one- or two-factor structures. Two studies demonstrated a one-factor structure using exploratory factor analysis (EFA) among patients with heart failure or gastrointestinal symptoms (Lee et al., 2014) and CFA among university students (Kim and Lee, 2019). Other two studies found two-factor structures using EFA (Park, 2017) or using both EFA and CFA (Shin et al., 2020) in general populations, and the patterns of clustering items into factors differed among them.

In general, CFA, which confirms a hypothesized factor structure by using a theory or empirical evidence, is known to provide more-compelling evidence for structural validity than EFA (Polit and Yang, 2016). When the construct to be measured has no theoretical rationale of dimensions, EFA is used to provide empirical evidence for the CFA measurement model. In this case, both EFA and CFA are used for structural validity, which is called a cross-validation approach. EGA has recently emerged as a powerful assessment tool for identifying the number of factors that underlie multivariate data in network psychometrics (Golino and Epskamp, 2017). In a simulation study, EGA demonstrated greater accuracy than traditional factor analysis methods (e.g., EFA) in estimating the number of latent factors (Golino et al., 2020). EGA was therefore used instead of EFA to assess the number of factor structures in the PHQ-9 in the present study, which demonstrated a one-factor structure. The one-factor structure has also been demonstrated using EGA in patients living with epilepsy (Sebera et al., 2020).

In this study, the empirically derived one-factor structure of the PHQ-9 using EGA had satisfactory structural validity in the CFA. The one-factor structure in the present study has been previously demonstrated among nationally representative general populations in Peru ($n = 30,449$) (Villarreal-Zegarra et al., 2019) and European countries (Ui, $n = 2,025$; Ireland, $n = 1,041$; Spain, $n = 1,949$; and Italy, $n = 1,048$) (Shevlin et al., 2022). Some researchers insisted that the number of factor structures was likely to be one when the PHQ-9 was applied to a more-heterogeneous sample (e.g., a general population) because the item variance would be greater and thus items would load on one factor

(Petersen et al., 2015). However, the PHQ-9 was demonstrated to have a two-factor structure in data representative of the United States population from the 2005–2016 NHANES (Patel et al., 2019). The inconsistency in the number of factors therefore might not only be explained by the hetero- or homogeneous characteristics of a sample; the era and culture of the society and environment in which people are currently living may also be considered as the possible source of this inconsistency. For example, the content of item 6 (“Trouble concentrating on things, such as reading the newspaper or watching television”) might not have been a problem when the PHQ-9 was developed in 2001 (Kroenke et al., 2001). The internet penetration rate is 97% (DataReportal, 2021) and the rate of smartphone use among adults is 95% (Pew Research Center, 2019) in South Korea. The rapid adoption of internet and device technologies has resulted in the daily newspaper utilization rate decreasing from 87.8% in 1993 to 44.6% in 2011 and 8.9% in 2021, and over-the-top (OTT) media services were introduced in 2013 in South Korea (Korea Press Foundation, 2021). In other words, most South Koreans do not read the daily newspapers, and the pattern of watching TV has been moving from terrestrial television broadcasting toward OTT media services that can be watched anytime, anywhere, and with any device. In these conditions, the content of item 6 may lead to biased responses. It is therefore recommended to adjust the phrasing of the item 6 (“...reading the newspaper or watching television”) to correspond to the current circumstances.

The omega coefficient indicated that the internal consistency of the PHQ-9 was satisfactory in this study. Cronbach's alpha has been dominantly used to assess the internal consistency of a self-reported questionnaire. However, this metric criticized due to the violation of tau-equivalence, and the omega coefficient has emerged as a new alternative (Taylor, 2021). The omega value of the PHQ-9 was also supported by studies involving a general population in Peru ($\omega = 0.87$) (Villarreal-Zegarra et al., 2019) and university students in Bangladesh ($\omega = 0.86$) (Rahman et al., 2022).

Measurement invariance across sociodemographic and medical-condition groups

The measurement invariance of the PHQ-9 across age groups was supported in the present study. This was congruent with studies involving general populations in other countries (Villarreal-Zegarra et al., 2019; Lamela et al., 2020). The consistency of this finding implies

that the depressive symptoms scored by the PHQ-9 can be meaningfully compared among age groups in a general population.

Invariance has been reported in the PHQ-9 across sexes in the general population of the United States (Patel et al., 2019), and across primary-care patients in Spain (González-Blanch et al., 2018) and in Germany (Petersen et al., 2015). In the present study, partial scalar invariance across sex was yielded by the noninvariant intercept of item 3 (“Trouble falling, staying asleep or sleeping too much”). However, the effect of the partial scalar invariance on comparing mean differences between groups is small and is not practically relevant (Schuler et al., 2018). Considering this, the PHQ-9 was able to yield invariance across sex groups with minimal risk of bias in the present study. Partial scalar invariance was also yielded across marital groups. This finding was consistent with those of other studies involving general populations in Portugal and Peru (Villarreal-Zegarra et al., 2019; Lamela et al., 2020).

The PHQ-9 has been demonstrated to be invariant across education-level groups in a general population and in primary-care patients (González-Blanch et al., 2018; Patel et al., 2019; Lamela et al., 2020). However, the PHQ-9 in the present study yielded a partial metric invariant model across four education-level groups after freeing (unconstraining) the factor loading of item 9 (“Thought that you would be better off dead, or of hurting yourself in some way”). That is, the meaning of the item 9 differed across the education-level groups. South Korea is well known as a country in which educational competition is very higher and educational achievement is very important (Kwak and Ickovics, 2010). Because people with lower education levels are more likely to experience financial constraints, and being poor is known to be a major reason for suicide in the country (Kim et al., 2010; Lee et al., 2017; Hong et al., 2021), the suicide rate has been ranked the highest among the Organization for Economic Cooperation and Development (OECD) countries (OECD, 2022). People with different education levels therefore seem to respond differently to item 9 about suicide ideation.

The noninvariance of item 9 in this study might have also occurred due to the potential for response bias. While other studies administered the PHQ-9 using an internet or paper-pencil mode (González-Blanch et al., 2018; Patel et al., 2019; Lamela et al., 2020), the PHQ-9 survey was performed in this study using face-to face interviews. Since item 9 is a very sensitive question, respondents might not answer it frankly in an interview. Another potential reason is the content of the item 9 itself. Controversy exists regarding the item because its content of self-harm is not part of the nine criteria for depressive symptoms from the DSM that had used in the development of the PHQ-9 (Kroenke et al., 2009; Wu et al., 2020). For this reason, the PHQ-8 omits item 9 from the PHQ-9 and can be utilized in a large general population, such as in the Behavioral Risk Factor Surveillance System survey in the United States (Kroenke et al., 2009). It is therefore recommended to use self-reported paper-and-pencil/internet modes rather than an interview mode for the PHQ-9, or to consider using the PHQ-8 with item 9 omitted if the PHQ-8 has satisfactory psychometric properties in a specific population.

The meaning of the PHQ-9 in the present study was equivalent for people with and without each medical condition (hypertension, diabetes, cancer, arthritis, asthma, and heart disease), which suggests that researchers and health professionals can use the PHQ-9 to reliably compare between groups. The measurement invariance across medical-condition groups has rarely been studied. It is therefore

further recommended to test such psychometric validation in various disease groups.

Limitations

The data analyzed in this study were collected using a cross-sectional design, and so the measurement invariance of the PHQ-9 could not be examined over time. It is therefore recommended to assess whether the one-factor model of the PHQ-9 is invariant over time. In contrast to the hypertension and diabetes groups, people with and without cancer, arthritis, asthma, and heart disease were determined using self-reported physical diagnoses that might induce bias in the diagnostic accuracy. The samples of the medical-condition groups (asthma and heart disease) were too small, and so their findings of invariance should be interpreted with caution. This was the secondary analysis of a large data set, so the other psychometric examinations (e.g., test-retest reliability, convergent validity, criterion validity, and responsiveness) were not available for the PHQ-9.

Conclusion

The one-factor model of the PHQ-9 confirmed in this study empirically supported its measurement invariance across various sociodemographic and medical-condition groups. In other words, the meaning of the PHQ-9 was similar to people across the groups. Therefore, the PHQ-9 can be reliably used to compare the severity of depressive symptoms across the groups in research and practice.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: data used in this study are available on the website Korea National Health & Nutrition Examination Survey at https://knhanes.kdca.go.kr/knhanes/sub03/sub03_02_05.do.

Ethics statement

The studies involving humans were approved by Institutional Review Board at Ajou University Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

E-HL: study conception and design, data analysis and interpretation, draft preparation, manuscript writing and editing, and approval of the submitted version of the manuscript. EK, and H-JK: data screening, literature review and manuscript editing, and approval of the submitted version of the manuscript. HL: data analysis and interpretation, manuscript editing, and approval of the submitted

version of the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The researchers would like to thank the respondents who participated in the national survey.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Alpizar, D., Plunkett, S. W., and Whaling, K. (2018). Reliability and validity of the 8-item patient health questionnaire for measuring depressive symptoms of Latino emerging adults. *J. Lat/ Psychol.* 6, 115–130. doi: 10.1037/lat0000087
- American Psychiatric Association. *Diagnosis and Statistical Manual of Mental Disorders DSM-IV-TR*. 4th Washington DC: American Psychiatric Association (2000).
- Arrieta, J., Aguerrebere, M., Raviola, G., Flores, H., Elliott, P., Espinosa, A., et al. (2017). Validity and utility of the patient health questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: a cross-sectional study. *J. Clin. Psychol.* 73, 1076–1090. doi: 10.1002/jclp.22390
- Boothroyd, L., Dagnan, D., and Muncer, S. (2019). PHQ-9: one factor or two? *Psychiatry Res.* 271, 532–534. doi: 10.1016/j.psychres.2018.12.048
- Byrne, B. M. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. 3rd New York: Routledge (2016).
- Carroll, H. A., Hook, K., Perez, O. F. R., Denckla, C., Vince, C. C., Ghebrehiwet, S., et al. (2020). Establishing reliability and validity for mental health screening instruments in resource-constrained settings: systematic review of the PHQ-9 and key recommendations. *Psychiatry Res.* 291:113236. doi: 10.1016/j.psychres.2020.113236
- Centers for Disease Control and Prevention. National health nutrition examination survey. (2020). Available at: <https://www.cdc.gov/nchs/nhanes/index.htm> (assessed October 25, 2022).
- Chen, F. F. (2007). Sensitivity of goodness of fit indexed to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834
- Chen, Y. J., and Tang, T. L. P. (2006). Attitude toward and propensity to engage in unethical behavior: measurement invariance across major among university students. *J. Bus. Ethics* 69, 77–93. doi: 10.1007/s10551-006-9069-6
- Christensen, A. P., and Golino, H. (2021a). On the equivalency of factor and network loadings. *Behav. Res. Methods* 53, 1563–1580. doi: 10.3758/s13428-020-01500-6
- Christensen, A. P., and Golino, H. (2021b). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: a Monte Carlo simulation and tutorial. *Psych* 3, 479–500. doi: 10.3390/psych3030032
- Collier, J. E. *Applied structural equation Modeling using AMOS: Basic to advanced techniques*. New York: Routledge (2020).
- DataReportal. Digital 2021: South Korea. (2021). Available at: <https://datareportal.com/reports/digital-2021-south-korea>. (accessed February 3, 2022).
- Davis, K. A. S., Coleman, J. R. L., Adams, M., Allen, N., Breen, G., Cullen, B., et al. (2020). Mental health in UK biobank – development, implementation and results from an online questionnaire completed by 157366 participants: a reanalysis. *BJPsych Open* 6:e18. doi: 10.1192/bjo.2019.100
- El-Den, S., Chen, T. F., Gan, Y. L., Wong, E., and O'Reilly, C. L. (2018). The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. *J. Affect. Disord.* 225, 503–522. doi: 10.1016/j.jad.2017.08.060
- Golino, H., and Christensen, A. P. EGAnet: Exploratory graph analysis- a framework for estimating the number of dimensions in multivariate data using network psychometrics. R package version 1.2.0. (2022).
- Golino, H. F., and Epskamp, S. (2017). Exploratory graph analysis: a new approach for estimating the number of dimensions in psychological research. *PLoS One* 12:e0174035. doi: 10.1371/journal.pone.0174035
- Golino, H., Lillard, A. S., Becker, I., and Christensen, A. P. (2021). Investigating the structure of the children's concentration and empathy scale using exploratory graph analysis. *PTAD* 2, 35–49. doi: 10.1027/2698-1866/a000008

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1217038/full#supplementary-material>

- Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., et al. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: a simulation and tutorial. *Psychol. Methods* 25, 292–320. doi: 10.1037/met0000255
- González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruiz-Rodríguez, P., Moriana, J. A., Limonero, J. T., et al. (2018). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS One* 13:e0193356. doi: 10.1371/journal.pone.0193356
- Greenberg, P. E., Fournier, A. A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S. H., et al. (2021). The economic burden of adults with major depressive disorder in the United States (2010 and 2018). *Pharmacoeconomics* 39, 653–665. doi: 10.1007/s40273-021-01019-4
- Hinz, A., Mehnert, A., Kocalevent, R. D., Brähler, E., Forkmann, T., Singer, S., et al. (2016). Assessment of depression severity with the PHQ-9 in cancer patients and in the general population. *BMC Psychiatry* 16:22. doi: 10.1186/s12888-016-0728-6
- Hong, M. W., Lee, J. H., Lee, K. W., Kim, S. B., and Kang, M. G. (2021). Risk factors for depressive symptoms in Korean adult stroke survivors: the Korea national health and nutrition examination survey IV–VII (2007–2018). *Int. J. Environ. Res. Public Health* 18:8178. doi: 10.3390/ijerph18158178
- Hong, J. W., Noh, J. H., and Kim, D. J. (2021). The prevalence of and factors associated with depressive symptoms in the Korean adults: the 2014 and 2016 Korea national health and nutrition examination survey. *Soc. Psychiatry Psychiatr. Epidemiol.* 56, 659–670. doi: 10.1007/s00127-020-01945-2
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* 6, 1–55. doi: 10.1080/10705519909540118
- Jang, I. (2021). Pre-hypertension and its determinants in healthy young adults: analysis of data from the Korean national health and nutrition examination survey VII. *Int. J. Environ. Res. Public Health* 18:9144. doi: 10.3390/ijerph18179144
- Keum, B. T., Miller, M. J., and Inkelas, K. K. (2018). Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. *Psychol. Assess.* 30, 1096–1106. doi: 10.1037/pas0000550
- Kim, M. H., Jung-Choi, K., Jun, H. J., and Kawachi, I. (2010). Socioeconomic inequalities in suicidal ideation, parasuicides, and completed suicides in South Korea. *Soc. Sci. Med.* 70, 1254–1261. doi: 10.1016/j.socscimed.2010.01.004
- Kim, Y. E., and Lee, B. (2019). The psychometric properties of the patient health questionnaire-9 in a sample of Korean university students. *Psychiatry Investig.* 16, 904–910. doi: 10.30773/pi.2019.0226
- Korea Centers for Disease Control and Prevention. Korea health statistics 2020: Korea National Health and nutrition examination survey (KNHANES VIII-2). (2022). Available at: <https://knhanes.kdca.go.kr> (accessed September 29, 2022).
- Korea Press Foundation. *Newspaper Industry in Korea 2021*. Seoul: Korea Press Foundation (2021).
- Kroenke, K., Spitzer, R. L., and Williams, J. B. W. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x
- Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., and Mokdad, A. H. (2009). The PHQ-8 as a measurement of current depression in the general population. *J. Affect. Disord.* 114, 163–173. doi: 10.1016/j.jad.2008.06.026
- Kwak, C. W., and Ickovics, J. R. (2010). Adolescent suicide in South Korea: risk factors and proposed multiple-dimensional solution. *Asian J. Psychiatr.* 43, 150–153. doi: 10.1016/j.ajp.2019.05.027

- Lamela, D., Soreira, C., Matos, P., and Morais, A. (2020). Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. *J. Affect. Disord.* 276, 220–233. doi: 10.1016/j.jad.2020.06.066
- Lee, E. J., Hall, L. A., and Moser, D. K. (2014). Psychometric properties of the patient health questionnaire-9 in patients with heart failure and gastrointestinal symptoms. *J. Nurs. Meas.* 22, 29E–40E. doi: 10.1891/1061-3749.22.2.29
- Lee, S. U., Oh, I. H., Jeon, H. J., and Roh, S. (2017). Suicide rates across income levels: retrospective cohort data on 1 million participants collected between 2003 and 2013 in South Korea. *J. Epidemiol.* 27, 258–264. doi: 10.1016/j.je.2016.06.008
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130
- McDonald, R. P. *Test Theory: A Unified Treatment*. Mahwah: Lawrence Erlbaum (1999).
- Mwangi, P., Nyongesa, M. K., Koot, H. M., Cuijpers, P., Newton, C. R., and Abubakar, A. (2020). Validation of a Swahili version of the 9-item patient health questionnaire (PHQ-9) among adults living with HIV compared to a community sample from Kilifi. *Kenya. J. Affect. Disord. Rep.* 1:100013. doi: 10.1016/j.jadr.2020.100013
- Nouwen, A., Deschênes, S. S., Balkhiyarova, Z., Albertorio-Díaz, J. R., Prokopenko, I., and Schmitz, N. (2021). Measurement invariance testing of the patient health questionnaire-9 (PHQ-9) across people with and without diabetes mellitus from the NHANES, EMHS and UK biobank datasets. *J. Affect. Disord.* 292, 311–318. doi: 10.1016/j.jad.2021.05.031
- OECD. *Suicide Rates*. (2022). OECD. Paris
- Park, K. Y. (2017). Reliability, validity and clinical usefulness of the Korean version of the patient health questionnaire-9 (PHQ-9). *Global Health Nurs.* 7, 71–78. doi: 10.35144/ghn.2017.7.2.71
- Patel, J. S., Oh, Y., Rand, K. L., Wu, W., Cyders, M. A., Kroenke, K., et al. (2019). Measurement invariance of the patient health questionnaire-9 (PHQ-9) depression screener in U.S. across sex, race/ethnicity, and educational level: NHANES 2005–2016. *Depress. Anxiety* 36, 813–823. doi: 10.1002/da.22940
- Petersen, J. J., Paulitsch, M. A., Hartig, J., Mergenthal, K., Gerlach, F. M., and Gensichen, J. (2015). Factor structure and measurement invariance of the patient health questionnaire-9 for female and male primary care patients with major depression in Germany. *J. Affect. Disord.* 170, 138–142. doi: 10.1016/j.jad.2014.08.053
- Pew Research Center. Smartphone ownership is growing rapidly around the world, but not always equally. (2019). Available at: <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally> (accessed February 3, 2022).
- Polit, D. F., and Yang, F. M.. *Measurement and the measurement of change*. Philadelphia: Wolters Kluwer (2016).
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., et al. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual. Life Res.* 27, 1147–1157. doi: 10.1007/s11366-018-1798-3
- Rahman, M. A., Dhira, T. A., Sarker, A. R., and Mehareen, J. (2022). Validity and reliability of the patient health questionnaire scale (PHQ-9) among university students of Bangladesh. *PLoS One* 17:e0269634. doi: 10.1371/journal.pone.0269634
- Schreiber, J. B. (2008). Core reporting practices in structural equation modeling. *Res. Social Adm. Pharm.* 4, 83–97. doi: 10.1016/j.sapharm.2007.04.003
- Schuler, M., Strohmayer, M., Mühlhig, S., Schwaighofer, B., Wittmann, M., Faller, H., et al. (2018). Assessment of depression before and after inpatient rehabilitation in COPD patients: psychometric properties of the German version of the patient health questionnaire (PHQ-9/PHQ-2). *J. Affect. Disord.* 232, 268–275. doi: 10.1016/j.jad.2018.02.037
- Sebera, F., Vissoci, J. R. N., Umwiringirwa, J., Teuwen, D. E., Boon, P. E., and Dedeken, P. (2020). Validity, reliability and cut-offs of the patient health questionnaire-9 as a screening tool for depression among patients living with epilepsy in Rwanda. *PLoS One* 15:e0234095. doi: 10.1371/journal.pone.0234095
- Shevlin, M., Butter, S., McBride, O., Murphy, J., Gibson-Miller, J., Hartman, T. K., et al. (2022). Measurement invariance of the patient health questionnaire (PHQ-9) and generalized anxiety disorder scale (GAD-7) across four European countries during the COVID-19 pandemic. *BMC Psychiatry* 22:154. doi: 10.1186/s12888-022-03787-5
- Shin, C., Ko, Y. H., An, H., Yoon, H. K., and Han, C. (2020). Normative data and psychometric properties of the patient health questionnaire-9 in a nationally representative Korean population. *BMC Psychiatry* 20:194. doi: 10.1186/s12888-020-02613-0
- Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Grossman, D. C., Baumann, L. C., Davidson, K. W., et al. (2016). Screening for depression in adults: US preventive services task force recommendation statement. *JAMA* 315, 380–387. doi: 10.1001/jama.2015.18392
- Stochl, J., Fried, E. I., Fritz, J., Croudace, T. J., Russo, D. A., Knight, C., et al. (2022). On dimensionality, measurement invariance, and suitability of sum scores for the PHQ-9 and the GAD-7. *Assessment* 29, 355–366. doi: 10.1177/1073191120976863
- Taylor, J. M. (2021). Coefficient omega. *J. Nurs. Educ.* 60, 429–430. doi: 10.3928/01484834-20210722-02
- Teymouri, A., Real, R., Gorbunova, A., Haghish, E. F., Andelic, N., Wilson, L., et al. (2020). Measurement invariance of assessments of depression (PHQ-9) and anxiety (GAS-7) across sex, strata and linguistic backgrounds in a European-wide sample of patients after traumatic brain injury. *J. Affect. Disord.* 262, 278–285. doi: 10.1016/j.jad.2019.10.035
- Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabé-Ortiz, A., Melendez-Torres, G. J., and Bazo-Alvarez, J. C. (2019). Valid group comparisons can be made with the patient health questionnaire (PHQ-9): a measurement invariance study across groups by demographic characteristics. *PLoS One* 14:20221717. doi: 10.1371/journal.pone.0221717
- World Health Organization. Depression (2021). Available at: https://www.who.int/health-topics/depression#tab=tab_1 (accessed December 21, 2022).
- Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., et al. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. *Psychol. Med.* 50, 1368–1380. doi: 10.1017/S0033291719001314