

Standardization of the Korean Version of the Patient Health Questionnaire-4 (PHQ-4)

Hyoun-Wook Kim^{1,*}, Cheolmin Shin^{2,*}, Seung-Hoon Lee³, Changsu Han²

¹Department of Psychiatry, Gwangju Jeonnam Regional Military Manpower Administration, Gwangju, ²Department of Psychiatry, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, ³Department of Psychiatry, Veterans Health Service Medical Center, Seoul, Korea

Objective: The Patient Health Questionnaire-4 (PHQ-4) has been used for screening owing to ease of use and brevity. In this study, we developed the Korean version of the PHQ-4 and tested its validity.

Methods: One hundred sixteen new adult outpatients at the Department of Psychiatry of the Korea University Ansan Hospital participated in the study. We simultaneously administered other depression/anxiety scales: the Hamilton Rating Scale for Depression, the Hamilton Anxiety Scale, the Beck Depression Inventory, and the Beck Anxiety Inventory.

Results: The mean PHQ-4 score was 6.52 (standard deviation = 3.45). Cronbach's α was 0.792, and the intraclass correlation coefficient of test and 2-week interval retest was 0.827 ($p < 0.01$). The Pearson correlation coefficients between the PHQ-4 total score and other depression/anxiety scales were all over 0.6. Confirmatory factorial analysis showed acceptable convergent validity and reliability but questionable discriminant validity for some model fit values.

Conclusion: The Korean version of the PHQ-4 has sufficient internal consistency, test-retest reliability, and construct validity, but its two-factor structure showed incompleteness. However, we suggest that it should be used as a brief screening measure for common psychiatric distress that warrants further detailed assessment, but not to separately assess the severity of depression and anxiety symptoms.

KEY WORDS: Patient Health Questionnaire-4; Standardization; Depression; Anxiety; Screening.

INTRODUCTION

Depressive disorders are among the most common psychiatric disorders and a major cause of disability [1]. These disorders account for 3.0% (2.2–3.8%) of global Disability-Adjusted Life Years (DALY), contributing to 16 million suicidal DALYs and 4 million ischemic heart disease DALYs per year [2]. In the United States, the economic cost of major depressive disorders rose by 21.5% over five years to \$210.5 billion in 2010. In Korea, the prevalence rate of major depression disorders has risen to 3.6% [3]. Anxiety disorders are also among the most common psychiatric disorders, with a global prevalence rate

of 28.8% [4], and a 9.1% rate in Korea specifically [3]. In the United States, the economic cost of anxiety disorders totals \$42.3 billion [5].

However, many patients afflicted with these diseases do not receive proper treatment. Only 2.2% of patients with symptoms of depression visit a doctor, and only 7.4% of them receive antidepressants [3]. A study of clinical practice guidelines for depression and the education of primary physicians showed that trained doctors were able to identify 39% of depressed patients [6]. Another study showed that 41% of patients with anxiety disorders were not receiving treatment [7]. One reason for this lack of treatment is time constraints; 90% of primary physicians said that more time is needed to diagnose depression compared to other diseases [8].

There are some measures used in the screening of depressive and anxiety disorders. The Patient Health Questionnaire (PHQ-9) measures the severity of depression. Each of the nine questions evaluates the symptoms of the

Received: March 17, 2020 / **Accepted:** March 26, 2020

Address for correspondence: Changsu Han

Department of Psychiatry, Korea University Ansan Hospital, Korea University College of Medicine, 123 Jeokgeum-ro, Danwon-gu, Ansan 15355, Korea

E-mail: hancs@korea.ac.kr

ORCID: <https://orcid.org/0000-0002-4021-8907>

*These authors contributed equally to this study as co-first authors.

© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diagnostic and Statistical Manual of Mental Disorders 4th edition (DSM-IV), the diagnostic standard for major depressive disorders [9]. The PHQ-2 is derived from the PHQ-9 and measures only two key symptoms of the diagnostic standard: depression and decreased interest. This reduces the time and effort required for the examination, which makes it suitable for screening [10]. The Generalized Anxiety Disorder Scale-7 (GAD-7) is a questionnaire that measures the severity of generalized anxiety disorders [11]. The GAD-2 consists of two core items from the GAD-7, and is also suitable for screening purposes, as it shows good sensitivity in diagnosing not only generalized anxiety disorders, but also panic disorder, social anxiety disorder, and post-traumatic stress disorder [7].

PHQ-4 combines the items from the PHQ-2 and GAD-2 [12]; therefore, it is a brief screening tool for assessing the key symptoms of depression and anxiety. As mentioned above, depressive and anxiety disorders show high prevalence rates and frequent comorbidity [13,14]. The degree of the resultant disability increases further in cases of comorbidity [15-17] and thus, the PHQ-4 is a useful tool for screening both disorders. The article that developed the PHQ-4 confirmed the construct and factor validities and the total scores were found to be related to the degree of functional disability [12]. In another study, the PHQ-4's validity in the general population was confirmed and normative data was obtained [18].

Due to cultural differences, the validity of screening questionnaires can be negatively impacted when translated into other languages. Therefore, it is necessary to develop a Korean version of the PHQ-4 and to verify its validity and reliability. If the Korean version of the PHQ-4 is developed through this study, it could be used as a brief and simple screening tool for depressive and anxiety disorders in a wide range of clinical environments, including psychiatry, primary care, and secondary/tertiary medical institutions. This further promotes public mental health practices.

METHODS

Study Design and Participants

The study was conducted between August 2017 and December 2018. The participants were new outpatients at the Department of Psychiatry of Korea University Ansan Hospital, aged 19 years or older. The study excluded pa-

tients whose cognitive abilities had deteriorated to such a degree that they would have had difficulty understanding and completing the research consent forms and questionnaires, or participating in interviews. The target sample size was determined to be 112; this assumed a 10% dropout rate, with an α level of 0.05 and power of 0.8, in accordance with the Cronbach's α figure derived from existing standardization studies [18,19].

Following the completion of a self-report questionnaire and an observer scale, the participants were examined and then retested two weeks later. One psychiatric resident doctor (HW Kim) completed all observer scales.

All participants signed an informed consent form and all research processes were conducted under the approval of the institutional review board (IRB) of the Korea University Ansan Hospital (IRB no. 2017AS0051).

Measures

Korean version of the PHQ-4

The Korean version of the PHQ-4 consists of the first two items of the Korean version of the PHQ-9 [20] and the first two items of the Korean version of the GAD-7. The items measure core symptoms of major depressive disorder and generalized anxiety disorder; items from the PHQ-9 ask about depression and anxiety, items from the GAD-7 ask about nervous feelings and uncontrollable worrying. All items are scored from 0 to 3 depending on the frequency patients have been bothered by the symptoms over the preceding two weeks. Our research team developed the Korean version of the GAD-7 through a consecutive process of translation and retranslation, until clinical psychiatrists confirmed that the translation fully reflected the meaning of the original GAD-7 and was suitable for Koreans. A paper concerning this development and the validation of the Korean version of the GAD-7 will be published soon.

Korean version of the Hamilton Rating Scale for Depression

The Hamilton Rating Scale for Depression (HAM-D) is considered the gold-standard for observer-rating scales for depression [21] and was initially composed of 21 items. However, the last four items were deemed to be less informative and lacked internal consistency, so the 17-item version is now widely used [22]. The HAM-D is a

semi-structured interview scale in which the observer evaluates the following: nine items (depressive mood, guilty feeling, suicidal ideation, work and activity, psychomotor retardation or agitation, mental anxiety, physical anxiety, and hypochondriasis) which are graded from 0 to 4 points; seven items (early, mid-term, late insomnia, gastrointestinal somatic symptoms, general somatic symptoms, sexual symptoms, and insight) which are graded from 0 to 2 points; and one item for weight loss graded from 0 to 3 points. The reliability and validity of the Korean version of the HAM-D has been demonstrated [23].

Korean version of the Hamilton Anxiety Rating Scale

The Hamilton Anxiety Scale (HAM-A) is a clinician-rating scale consisting of 14 questions that measure the severity of anxiety symptoms, with a total score that ranges from 0 to 56 [24]. The HAM-A measures both general psychic anxiety (agitation and psychological distress) and somatic anxiety (anxiety-related physical complaints).

Korean version of the Beck Depression Inventory

The Beck Depression Inventory (BDI) is a self-reporting measure developed to assess depressive symptoms and their severity [25]. The BDI-II was then developed to meet the diagnostic criteria of the DSM-IV [26]. It has 21 questions with a total score that ranges from 0 to 63. The validity of the Korean version of the BDI-II has been demonstrated [27].

Korean version of the Beck Anxiety Inventory

The Beck Anxiety Inventory (BAI) is a 21-item self-report questionnaire widely used as a simple and highly relevant measure that assesses anxiety symptoms and their severity [28]. The Korean version of the BAI was translated by Kwon and Oei [29].

Korean version of the Mini-International Neuropsychiatric Interview

The Mini-International Neuropsychological Interview (MINI) is a structured interview that diagnoses psychiatric disorders based on the DSM-IV and the International Classification of Diseases version 10. Minimal training is required to administer this simple, brief, and effective (15 minute) interview [30]. The validity of the Korean version

of the MINI has been demonstrated [31].

Data Analysis

In this study, Cronbach's α was used to analyze the internal consistency of the Korean version of the PHQ-4. In addition, the intraclass correlation coefficient was obtained from the tests conducted at the two-week interval to explore the test-retest reliability.

To determine construct validity, we obtained the correlation between the total PHQ-4 score and each score from the following scale measures: HAM-D, HAM-A, BDI, and BAI. The correlation between the PHQ-4 depression/anxiety subtotal score and the above scale measures was also obtained.

Factorial validity was examined using exploratory factor analysis and confirmatory factor analysis. For confirmatory factor analysis, we constructed a structural equation model (SEM) that holds two factors, depression and anxiety, which was regressed from first two items and last two items of the PHQ-4. The convergent validity and reliability of the model was examined through composite validity (CR) and the average variance extracted (AVE). The CR should be greater than 0.7, and the AVE should be greater than 0.5 [32]. Discriminant validity was also examined by confirming whether the maximum shared variance (MSV) was less than the AVE. The model fit was evaluated using a normed chi-square (CMIN/df), the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the Tucker–Lewis index (TLI), and the comparative-fit-index (CFI). The respective threshold of each measurement is as follows: < 3 , < 0.05 , < 0.09 , > 0.95 , and > 0.95 [33]. All statistical analyses were conducted using IBM SPSS Statistical Package for the Social Sciences ver. 23.0 (IBM Co., Armonk, NY, USA) and IBM SPSS Amos ver. 23.0 (IBM Co., Armonk, NY, USA).

RESULTS

Sample characteristics

One hundred sixteen individuals participated in the study. The mean of the PHQ-4 total scores was 6.52 with a standard deviation (SD) of 3.45. The sociodemographic characteristics of the sample are described in Table 1. ANOVA analysis of the PHQ-4 scores showed significant differences in those aged 51 or above compared to other

Table 1. Sociodemographic characteristics and PHQ-4 scores of the study sample

Characteristic	Number	%	PHQ-4 score		<i>p</i> value	
			Mean	SD		
Sex	Male	65	56.0	6.80	3.725	0.336
	Female	51	44.0	6.16	3.064	
Age group	19–30	65	56.0	7.33	3.371	0.014*
	31–40	13	11.2	6.31	2.926	
	41–50	15	12.9	6.29	3.496	
	51–60	16	13.8	4.87	3.357	
	Over 61	7	6.0	3.57	3.047	
Marital status	Married	36	31.0	5.00	3.618	0.004*
	Divorced/separated	7	6.0	7.86	1.864	
	Unmarried	70	60.3	7.26	3.257	
	Widowed	3	2.6	3.67	2.082	
Employment status	Employed	48	41.4	6.19	3.518	0.393
	Unemployed	68	58.6	6.76	3.415	
Region	Metropolitan	9	7.8	5.56	3.779	0.422
	Suburban	106	91.4	6.57	3.432	
	Rural	1	0.9	10.00	-	
Religion	None	83	71.6	6.74	3.255	0.257
	Christian	17	14.7	6.18	3.762	
	Catholic	3	2.6	2.00	1.414	
	Buddhist	10	8.6	6.20	4.237	
Economic status	High	-	-	-	-	< 0.01*
	Moderate-high	48	41.4	5.28	3.308	
	Moderate-low	43	37.1	6.49	3.203	
	Low	23	19.8	8.83	2.980	

PHQ-4, The Patient Health Questionnaire-4; SD, standard deviation.

**p* < 0.05.

Table 2. Psychiatric diagnosis and total/subtotal PHQ-4 score of the study sample

Psychiatric diagnosis	Number	%	PHQ-4		PHQ-2		GAD-2	
			Mean	SD	Mean	SD	Mean	SD
Major depressive disorder	67	57.8	7.04	3.12	3.75	1.82	3.30	1.84
Bipolar disorder	2	1.7	8.50	0.71	5.00	0.00	3.00	1.00
Dysthymia	2	1.7	9.00	4.24	5.00	1.41	4.00	2.83
Generalized anxiety disorder	2	1.7	4.50	0.71	2.33	0.58	2.50	0.71
Panic disorder	32	27.6	6.91	3.70	3.39	2.18	3.42	2.00
Social/specific phobia	4	3.4	5.50	2.38	2.50	1.73	3.00	0.82
Obsessive-Compulsive disorder	2	1.7	8.50	0.71	2.50	0.71	6.00	0.00
Somatization disorder	1	0.9	9.00	0.00	3.00	0.00	6.00	0.00
Adjustment disorder	15	12.9	6.07	4.48	3.47	2.20	2.60	2.50
Alcohol dependence	2	1.7	7.00	7.07	4.00	2.83	3.00	4.24
Psychotic disorder	3	2.6	4.00	1.00	2.00	0.00	2.00	1.00

PHQ-4, The Patient Health Questionnaire-4; PHQ-2, The Patient Health Questionnaire-2; GAD-2, The Generalized Anxiety Disorder Scale-2; SD, standard deviation.

age groups, in those who were divorced or separated relative to other marital statuses, and in the low income group relative to those with other income levels.

The psychiatric diagnoses of the study sample are described in Table 2. Those who were diagnosed with two

or more disorders were assigned to both categories. The sample sizes were too small for categorization into specific disorders except major depressive disorder, panic disorder, and adjustment disorder. In general, the depression subtotal score (PHQ-2 score) relative to the anxiety sub-

total score (GAD-2) score tended to be higher in anxiety disorders, and vice versa.

Reliability

The Cronbach’s α value for the Korean version of the PHQ-4 was measured at 0.792, which suggests acceptable internal consistency [34]. As described in Table 3, all items showed significant associations with the total PHQ-4 score. When each item was removed, the Cronbach’s α value did not increase.

The intraclass correlation coefficient for test-retest values was excellent at 0.827, with a p value less than 0.01 [35]. The average of the first PHQ-4 scores was 6.54 with a SD of 3.44; whereas the average of the second PHQ-4 scores was 5.50 with a SD of 3.94.

Construct validity

The correlation analysis to measure structural validity is described in Table 4. It was observed that all scores and subscales of the PHQ-4 were sufficiently correlated with BDI, BAI, HAM-D, HAM-A, and had a Pearson correlation coefficient of 0.5 or more. However, the total PHQ-4 score showed a higher correlation to BDI and HAM-D than the PHQ-4 depression subscale, and was also higher

than the correlation between the PHQ-4 anxiety subscale and the BAI and HAM-A.

Factorial validity

The exploratory factor analysis demonstrated a 1-factor structure of the PHQ-4. The factor loadings for items 1 to 4 were 0.81, 0.80, 0.79, and 0.74, respectively.

In Figure 1A, the SEM for confirmatory factor analysis includes two factors, depression and anxiety, the standardized regression weights for each questionnaire item, and the covariance between them. Standardized regression weights were all above 0.5, which means that the two factors show sufficient explanatory power for each item. However, the covariance between the depression and anxiety factors was also high, suggesting that these two factors are not sufficiently independent. The AVE and CR for depression were 0.566 and 0.721, and 0.605 and 0.754 for anxiety, these all sufficiently indicate convergent validity and reliability. However, the MSV (0.585) was greater than the AVE for depression (0.566), confirm-

Table 3. Corrected item-total correlations and Cronbach’s α if an item is deleted

Item	Corrected item-total correlation	Cronbach’s α if an item deleted
1	0.543	0.772
2	0.643	0.721
3	0.631	0.729
4	0.601	0.742

Table 4. Correlation between PHQ-4 and concurrent validity measures

Scale	PHQ-4 total score	Depression score	Anxiety score
BDI	0.762**	0.656**	0.678**
HAM-D_17	0.720**	0.654**	0.615**
BAI	0.728**	0.567**	0.716**
HAM-A	0.680**	0.548**	0.647**

PHQ-4, The Patient Health Questionnaire-4; BDI, The Beck Depression Inventory; HAM-D, The Hamilton Rating Scale for Depression; BAI, The Beck Anxiety Inventory; HAM-A, The Hamilton Anxiety Scale.

** $p < 0.01$.

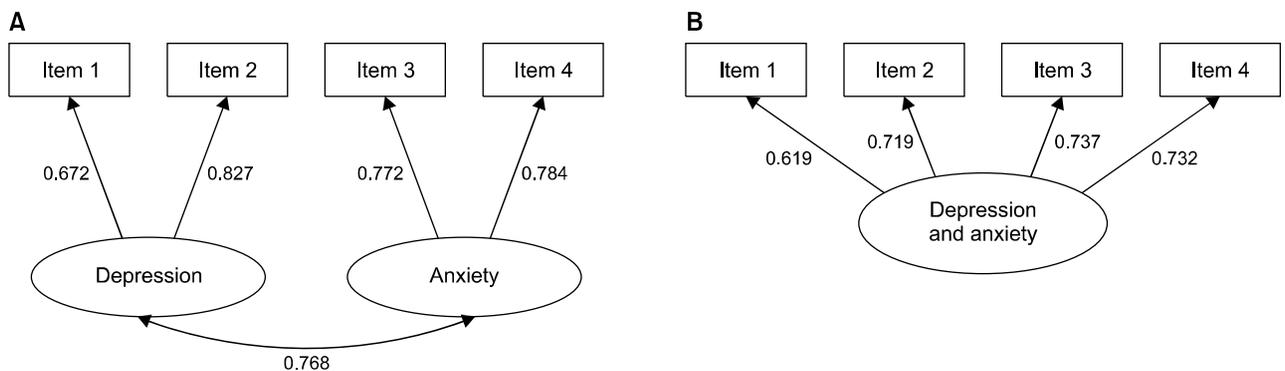


Fig. 1. Structural equation model of (A) 2-factor and (B) 1-factor model. Arrows from factors to items represent factor loading. Arrow between factors represent correlation.

ing that the model has insufficient discriminant validity. The measurement of model fit is as follows: CMIN/df = 3.237; RMSEA = 0.141; SRMR = 0.023, TLI = 0.901, and CFI = 0.983. The SRMR and CFI values satisfied the threshold, otherwise the CMIN/df, RMSEA, and TLI did not satisfy but were close to the threshold.

After recognizing the result of exploratory factor analysis, we then constructed a 1-factor model (Fig. 1B) and tested whether it better explains the structure of the PHQ-4. The AVE was 0.459, which is under threshold, and the CR was 0.796. The following values were obtained for model fit: CMIN/df, 6.353; RMSEA, 0.219; SRMR, 0.054; TLI, 0.762; and CFI, 0.921. As observed, all values were unfavorable compared to the 2-factor model, and only the SRMR satisfied the threshold.

DISCUSSION

This study has several main findings. First, the Korean version of the PHQ-4 was found to be a valid and reliable instrument for screening depression and anxiety. Second, this is the first study that standardized the PHQ-4 in the psychiatric unit. Unlike studies on the general population or a primary care clinical sample, the 2-factor structure of the PHQ-4 was incomplete.

The construct validity of the PHQ-4 showed significance supported by high Pearson correlation coefficients with existing depression/anxiety scales. In the analysis of the correlations between the PHQ-4 and other existing depression/anxiety scales, the PHQ-4 total score was observed to have a higher correlation than the PHQ-4 depression/anxiety subscores. The high comorbidity of depressive and anxiety disorders [36] may have an effect on this finding. With both depression/anxiety subscores (synonymous with PHQ-2 and GAD-2 each) having cut points of 3 or greater, as suggested by previous studies [7,10], 50 (44.2%) patients have both depression and anxiety, 22 (19.5%) have only depression, 12 (10.6%) have anxiety only, and 29 (25.7%) have neither depression nor anxiety. Among the patients with depression, 69.4% have comorbid anxiety, and among the patients with anxiety, 80.6% have comorbid depression. The comorbidity of depression and anxiety in this study was higher than the previous PHQ-4 study conducted on primary-care clinical sample [12].

Using confirmatory factor analysis, the correlation of

the depression and the anxiety factors was high at 0.768, and the discriminant validity was insufficient. Several model fit values also suggest that the two-factor model may not be perfectly suitable. The incompleteness of the 2-factor structure found in this study was inconsistent with previous PHQ-4 studies, and this may be attributable to the sample characteristics. In this study, subjects were recruited from the psychiatric department, whereas other studies recruited from the general population or primary-care sites. The mean score of the PHQ-4 in this study was 6.52 (SD 3.45), and it is significantly higher than in the German general population (1.76, SD 2.06) [18], in the Columbian general population (1.27, SD 2.01) [37], in undergraduate students from the United States (2.98) [38], and in a primary-care site sample from the United States (2.5, SD 2.8) [12]. The high symptom burden and aforementioned high comorbidity of depression and anxiety would have blurred the distinction between the two factors.

The reliability test demonstrated that the Korean version of the PHQ-4 is a reliable instrument. The internal consistency of the Korean version of the PHQ-4 was found to have a sufficient value of 0.792. The value was slightly lower than previous PHQ-4 standardization studies ranging from 0.81 to 0.85 [12,18,37,38], but the value was close to them. This study is the first that reports the test-retest reliability of the PHQ-4. Although the treatment between 2 weeks of test-retest interval would have negatively affected it, the intraclass correlation coefficient representing test-retest reliability showed an excellent value of 0.827.

The usage of standard diagnostic interview and the demonstration of test-retest reliability are the strengths of the present study. However, the biased study sample is a limitation. Löwe *et al.* [18] recommended considering PHQ-4 scores of 6 (percentile 95.7%) or greater as “yellow flags”, but 54.9% in this study sample correspond to “yellow flags”. Since the recruiting site was the psychiatric department of the university hospital, more severe patients would have visited compared to a primary-care clinic. Other diagnoses except for major depressive disorder (57.6%), panic disorder (27.6%), and adjustment disorder (9.0%), were all below 5%.

In conclusion, the Korean version of the PHQ-4 showed sufficient internal consistency, test-retest reliability, and construct validity. However, its 2-factor struc-

ture was incomplete in a psychiatric clinic setting characterized by high psychiatric prevalence and comorbidity. Therefore, we suggest that it should be used as brief screening measure for general psychiatric distress that warrants further detailed assessment, but not to separately assess the severity of depression and anxiety symptoms in a psychiatric clinic setting. To investigate the normative data and to confirm the factor structure of the Korean version of the PHQ-4 in settings with relatively low prevalence and comorbidity, future studies should be done in the general population.

■ Acknowledgments

This study was supported by a Korea University grant (K2008501).

■ Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

■ Author Contributions

Conceptualization: Changsu Han. Data acquisition: Hyoun-Wook Kim, Cheolmin Shin, Seung-Hoon Lee. Formal analysis: Hyoun-Wook Kim, Cheolmin Shin. Supervision: Changsu Han. Writing—original draft: Hyoun-Wook Kim, Cheolmin Shin. Writing—review & editing: Cheolmin Shin, Seung-Hoon Lee. All authors read and approved the final manuscript.

■ ORCID

Hyoun-Wook Kim <https://orcid.org/0000-0002-7942-6025>
 Cheolmin Shin <https://orcid.org/0000-0002-8232-2921>
 Seung-Hoon Lee <https://orcid.org/0000-0001-5341-0933>
 Changsu Han <https://orcid.org/0000-0002-4021-8907>

REFERENCES

- Mathers CD, Loncar D. *Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med* 2006; 3:e442.
- Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. *Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med* 2013;10:e1001547.
- Ohayon MM, Hong SC. *Prevalence of major depressive disorder in the general population of South Korea. J Psychiatr Res* 2006;40:30-36.
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. *Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry* 2005;62:593-602.
- Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. *The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry* 2015;76:155-162.
- Thompson C, Kinmonth AL, Stevens L, Peveler RC, Stevens A, Ostler KJ, et al. *Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. Lancet* 2000;355:185-191.
- Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. *Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med* 2007;146:317-325.
- Wittchen HU, Pittrow D. *Prevalence, recognition and management of depression in primary care in Germany: the Depression 2000 study. Hum Psychopharmacol* 2002;17 Suppl 1:S1-S11.
- Kroenke K, Spitzer RL, Williams JB. *The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med* 2001;16: 606-613.
- Kroenke K, Spitzer RL, Williams JB. *The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care* 2003;41:1284-1292.
- Spitzer RL, Kroenke K, Williams JB, Löwe B. *A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med* 2006;166:1092-1097.
- Kroenke K, Spitzer RL, Williams JB, Löwe B. *An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics* 2009;50:613-621.
- Olfson M, Shea S, Feder A, Fuentes M, Nomura Y, Gameraoff M, et al. *Prevalence of anxiety, depression, and substance use disorders in an urban general medicine practice. Arch Fam Med* 2000;9:876-893.
- Schonfeld WH, Verboncoeur CJ, Fifer SK, Lipschutz RC, Lubeck DP, Buesching DP. *The functioning and well-being of patients with unrecognized anxiety disorders and major depressive disorder. J Affect Disord* 1997;43:105-119.
- Olfson M, Fireman B, Weissman MM, Leon AC, Sheehan DV, Kathol RG, et al. *Mental disorders and disability among patients in a primary care group practice. Am J Psychiatry* 1997;154:1734-1740.
- Andrews G, Henderson S, Hall W. *Prevalence, comorbidity, disability and service utilisation. Overview of the Australian National Mental Health Survey. Br J Psychiatry* 2001;178: 145-153.
- Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. *Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl* 2004;(420):38-46.

18. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *J Affect Disord* 2010;122:86-95.
19. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. *J Educ Behav Stat* 2002;27:335-340.
20. Han C, Jo SA, Kwak JH, Pae CU, Steffens D, Jo I, et al. Validation of the Patient Health Questionnaire-9 Korean version in the elderly population: the Ansan Geriatric study. *Compr Psychiatry* 2008;49:218-223.
21. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
22. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278-296.
23. Yi JS, Bae SO, Ahn YM, Park DB, Noh KS, Shin HK, et al. Validity and reliability of the Korean version of the Hamilton Depression Rating Scale (K-HDRS). *J Korean Neuropsychiatr Assoc* 2005;44:456-465.
24. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959;32:50-55.
25. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561-571.
26. Beck AT, Steer RA, Brown GK. *BDI-II, Beck depression inventory: manual*. San Antonio:Psychological Corporation;1996.
27. Sung H, Kim J, Park Y, Bai D, Lee S, Ahn H. A study on the reliability and the validity of Korean version of the Beck Depression Inventory-II (BDI-II). *J Korean Soc Biol Ther Psychiatry* 2008;14:201-212.
28. Beck AT, Steer RA. *Beck anxiety inventory: manual*. San Antonio:Psychological Corporation;1990.
29. Kwon SM, Oei TP. Differential causal roles of dysfunctional attitudes and automatic thoughts in depression. *Cognit Ther Res* 1992;16:309-328.
30. Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry* 1997;12:224-231.
31. Yoo SW, Kim YS, Noh JS, Oh KS, Kim CH, Namkoong K, et al. Validity of Korean version of the mini-international neuropsychiatric interview. *Anxiety Mood* 2006;2:50-55.
32. Hair J, Black W, Babin B, Anderson RE. *Multivariate data analysis*. New Jersey:Pearson Prentice Hall;2010.
33. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1-55.
34. DeVellis RF. *Scale development: theory and applications*. Thousand Oaks:Sage Publications;2016.
35. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284-290.
36. Gorman JM. Comorbid depression and anxiety spectrum disorders. *Depress Anxiety* 1996;4:160-168.
37. Kocalevent RD, Finck C, Jimenez-Leal W, Sautier L, Hinz A. Standardization of the Colombian version of the PHQ-4 in the general population. *BMC Psychiatry* 2014;14:205.
38. Khubchandani J, Brey R, Kotecki J, Kleinfelder J, Anderson J. The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Arch Psychiatr Nurs* 2016;30:457-462.