



Comparison of the Usefulness of the PHQ-8 and PHQ-9 for Screening for Major Depressive Disorder: Analysis of Psychiatric Outpatient Data

Cheolmin Shin, Seung-Hoon Lee, Kyu-Man Han, Ho-Kyoung Yoon, and Changsu Han[✉]

Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea

Objective This study aimed to demonstrate that the Patient Health Questionnaire (PHQ)-8 is not less useful than the PHQ-9 as a screening test for major depressive disorder (MDD).

Methods We performed a retrospective analysis of 567 patients in psychiatric outpatient units. The Mini International Neuropsychiatric Interview was used to diagnose MDD. We derived the validity and reliability of the PHQ-8 and PHQ-9. To evaluate the ability of the PHQ-8 and PHQ-9 to discriminate MDD, we drew receiver operating characteristic (ROC) curves and compared the areas under the curves (AUCs).

Results Of the 567 participants, 207 (36.5%) were diagnosed with MDD. Cronbach's α s for the PHQ-8 and PHQ-9 were 0.892 and 0.876, respectively. Similar to the PHQ-9, the PHQ-8 was also associated with scores on the Hamilton Depression Rating Scale in a correlation analysis. When we drew ROC curves for the PHQ-8 and PHQ-9, there was no statistically significant difference in the AUCs. With a cutoff score of 10, the PHQ-8 showed a sensitivity of 58.3%, specificity of 83.1%, positive predictive value of 53.4%, and negative predictive value of 85.7%.

Conclusion In a psychiatric outpatient sample, the PHQ-8 was as useful as the PHQ-9 for MDD screening.

Psychiatry Investig 2019;16(4):300-305

Key Words Depression, Major depressive disorder, PHQ-8, PHQ-9.

INTRODUCTION

Major depressive disorder (MDD) is a serious mental illness of major clinical importance that reduces quality of life and increases suicide rates.^{1,2} However, when properly detected, there are treatment methods that are effective at improving the condition. For this reason, considerable importance has been placed on screening tests to detect MDD, and there have been efforts to screen for MDD using various instruments. These screening tests are useful not only for the primary purpose of identifying individuals with MDD but also for ascertaining the prevalence of MDD within certain populations.^{3,4}

The Patient Health Questionnaire-9 (PHQ-9), a self-report questionnaire for the diagnosis and assessment of depression, includes nine items from the diagnostic criteria for MDD in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), with each item scored from 0 points, if absent, to 3 points, if severe.⁵ The PHQ-9 has been validated as a screening test by various studies not only in the general population but also in primary care settings and in specific disease populations.^{6,7} Among the items in the PHQ-9, Item 9 asks "How often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?" and allows for responses from "Not at all" to "Nearly every day." The presence of Item 9, asking about suicidal ideation, has been highlighted as a potential problem. Specifically, when the test is implemented in the mass population, even if a score of 1 or higher is indicated for a PHQ-9 item, it is difficult to provide appropriate coping services, including psychiatric evaluation. Therefore, in the United States, the PHQ-8, which is the same as the PHQ-9 but without Item 9, has been standardized and introduced, and is being used in studies on MDD

Received: December 5, 2018 Revised: January 21, 2019

Accepted: February 1, 2019

✉ **Correspondence:** Changsu Han, MD, PhD, MHSc

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

Tel: +82-31-412-5140, Fax: +82-31-412-5144, E-mail: hancs@korea.ac.kr

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://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.

prevalence in large populations.⁸

Likewise, in Korea, the PHQ-8 may be a more appropriate tool compared to the PHQ-9 for MDD screening. In Korea, the PHQ-9 has been introduced and used as an instrument to investigate the prevalence of depression in a national survey known as the Korea National Health and Nutrition Examination Survey, which is conducted by a survey interviewer.⁹ However, if the question regarding suicidal ideation (i.e., Item 9 of the PHQ-9) is answered, there is the problem that it is difficult to provide proper psychiatric intervention because the survey interviewers are not trained as mental health professionals.

Indeed, within Item 9, the thought that death would be preferable is a relatively passive concept that does not necessarily lead to suicide, while the thought of hurting oneself is a more active suicidal concept. However, in a medical environment, most patients who endorsed Item 9 only agreed with the passive ideation that “they would be better off dead”.^{10,11} In a study of 1,022 coronary artery disease patients, of those who scored 1 or higher on Item 9 of the PHQ-9, only 19.8% demonstrated actual suicidal ideation in an assessment using the Computerized Diagnostic Interview Schedule (C-DIS), and only 8.1% had a suicide plan.¹¹ In another study of cancer patients, an additional structured questionnaire on suicidal ideation was administered among patients who scored at least 1 point on Item 9, and only one-third of them responded that they actually had suicidal ideation.¹⁰ Another one-third denied having suicidal ideation, and the remaining one-third responded that they only had the passive thought that they would be better off dead; thus, it was concluded that a score of 1 point or higher on Item 9 does not always match with suicidal ideation. It therefore remains unclear what the PHQ-9’s Item 9 is evaluating, and since only a small percentage of patients who passively think about death or self-harm actually develop active self-aggressiveness, it has been proposed that the PHQ-8, which omits Item 9, would be a better alternative.

The PHQ-8 has been proven to be a valid and reliable tool to screen for depression in several populations despite the omission of one item from the PHQ-9.¹²⁻¹⁴ The criterion validity of the PHQ-8 has been demonstrated in adult community samples.^{13,14} The construct validity of the PHQ-8 was also supported in a population-based study in which MDD symptoms were recognized.¹³ In a study of U.S. military personnel, the PHQ-8 showed almost the same ability to detect depression when compared to the PHQ-9.¹⁵

The prevalence of suicidal ideation and the suicide rate in Korea are relatively high,^{16,17} and so it is not clear whether MDD can be identified as precisely with the PHQ-8 as with the PHQ-9, and it is possible that its validity as a screening test for MDD may be lower than that of the PHQ-9 itself.

Hence, in this study, we aimed to investigate whether the

PHQ-8 shows appropriate reliability and validity for use as a screening instrument for MDD in a Korean sample. In addition, we aimed to test its screening ability for MDD compared to the PHQ-9 and to demonstrate that the PHQ-8 is no less useful than the PHQ-9 as a screening test.

METHODS

Study sample

The study subjects were individuals who visited a psychiatric department at a university hospital in the Republic of Korea from January 2012 to December 2017 and were referred for psychological assessment. The inclusion criteria were as follows: 1) patients aged 19 or older, 2) new outpatients at the first psychiatric examination, and 3) patients who were able to read and write Korean. The exclusion criteria were as follows: 1) patients with cognitive impairment that disabled them to answer the questionnaires appropriately and 2) patients with underlying medical or surgical condition that could affect study evaluation. The data regarding these subjects consisted of hospital records during this period, which were analyzed retrospectively. This study focused on psychiatric outpatient treatment at a general hospital that provides outpatient services, inpatient services, and community mental health services.

The Korean adaptation of the PHQ-9 was administered to patients during waiting time as a routine check. The Mini-International Neuropsychiatric Interview (MINI) and the Hamilton Depression Rating Scale (HAM-D) were administered by a clinical psychologist, when patients first visited the psychiatric outpatient unit. This study was approved by the Institutional Review Board of Korea University Ansan Hospital (2018AS0197). This study was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

Diagnosis of MDD

MDD was diagnosed using the MINI, administered by a clinical psychologist. The MINI is a structured interview tool to diagnose psychiatric disease based on the DSM-IV-TR and the International Classification of Diseases-10 (ICD-10). In Korea, the Korean MINI (K-MINI) was standardized by Yoo et al.¹⁸

Measurements

Depression was evaluated using the PHQ-9, which is a self-report scale of depressive symptoms. The PHQ-9 consists of nine items reflecting almost exactly the nine diagnostic criteria for MDD in the DSM-IV. The nine items in the PHQ-9 ask about the frequency of depressive symptoms over the previous two weeks, and each is scored as 0 points for “Not at all,”

1 point for “Several days,” 2 points for “More than half the days,” or 3 points for “Nearly every day.” Thus, the highest possible total score is 27 points. The Korean PHQ-9 was standardized in the general elderly population by Han et al.¹⁹ and in a primary care setting by Choi et al.²⁰

Item 9 in the PHQ-9 asks, “Over the last 2 weeks how often have you been bothered by this problem: thoughts that you would be better off dead or hurting yourself in some way?” The PHQ-8, which has been used in numerous previous studies,^{11,13,21} excludes Item 9 from the PHQ-9 but retains the other eight items unchanged.

The extent of depressive symptoms was evaluated by a clinical psychologist using the HAMD, which is a clinician-administered scale of depressive symptoms.²² It was originally developed for measuring severity in patients already diagnosed with MDD, but its uses have since expanded, including in research to evaluate the effects of treatment, and it is currently considered the standard for observer rating scales for depression. The original scale consisted of 21 items, but the four items regarding diurnal variation, depersonalization-derealization, paranoid symptoms, and obsessive-compulsive symptoms, respectively, are not only rare in patients with depression but were also found to reduce internal consistency. Therefore, the 17-item version, which omits these items, is currently the most widely used version.²³ In this study, we used the Korean adaptation of the 17-item version of the HAMD, standardized by Yi et al.²⁴

Statistics

Among the sociodemographic variables, categorical variables were analyzed using the chi-square test and continuous variables were analyzed using the Student's *t*-test. In order to calculate the internal consistency of the PHQ-8 and PHQ-9, we used Cronbach's α . Through correlation analysis, we obtained Spearman coefficients for the HAMD with the PHQ-8 and PHQ-9. The above analyses were performed using SPSS version 20 (IBM Corp., Armonk, NY, USA). To derive the optimal cutoff points for the PHQ-8 and PHQ-9, we performed receiver operating characteristic (ROC) analysis. In order to compare the respective abilities of the PHQ-8 and PHQ-9 to diagnose MDD, we compared the areas under the ROC curves (AUCs). All statistical analysis was conducted using MedCalc Software bvba (Ostend, Belgium; <http://www.medcalc.org>; 2017). In all tests, statistical significance was defined as a *p*-value <0.05.

RESULTS

Sociodemographic data

A total of 567 eligible patients were registered. Table 1 shows

Table 1. Sociodemographic data and characteristics of study subjects

| Characteristics | N (%) |
|----------------------------|------------|
| Age | |
| 19–34 | 123 (21.7) |
| 35–60 | 288 (50.7) |
| 61+ | 157 (27.6) |
| Gender | |
| Male | 192 (33.8) |
| Female | 376 (66.2) |
| Education | |
| ≤6 | 126 (22.2) |
| >6, ≤12 | 279 (49.1) |
| >12 | 163 (28.7) |
| Marital status | |
| Married | 350 (61.6) |
| Separated/divorced/widowed | 98 (17.3) |
| Unmarried | 120 (21.1) |
| Occupational status | |
| Work | 227 (40.0) |
| Retired | 83 (14.6) |
| Homemaker | 134 (23.6) |
| Student | 33 (5.8) |
| No job | 91 (16.0) |

Data are expressed as number (%)

sociodemographic characteristics of subjects. The mean age was 49.6 years (SD=16.8), and there were 376 women (66.2%). The most common level of education was “≥6 years and ≤12 years” (n=279, 49.1%), most common occupational status was “employed” (n=227, 40.0%) and most common marital status was “married” (n=350, 61.6%).

The prevalence of all diagnoses was as follows: MDD (n=208, 36.6%), dysthymia (n=12, 2.1%), bipolar disorder with hypomanic or manic episodes (n=10, 1.8%), anxiety disorders (n=49, 8.6%), posttraumatic stress disorder (n=3, 0.5%), adjustment disorder (n=14, 2.5%), alcohol dependence (n=11, 1.9%), psychotic disorder (n=18, 3.2%), obsessive-compulsive disorder (n=5, 0.9%) somatization disorder (n=11, 1.9%), and psychiatric disorders not otherwise specified included unspecified depressive disorder, impulse control disorder, minor neurocognitive disorder, and sleep-related disorders (n=226, 39.6%).

Reliability and validity of the PHQ-8 and PHQ-9

Cronbach's α for the PHQ-8 was 0.88 and that for the PHQ-9 was 0.89. There was no statistically significant difference between the PHQ-8 and PHQ-9 in terms of internal consistency.

The sample in our study demonstrated sufficient convergent validity for the PHQ-8 as well as the PHQ-9 (Table 2). The analysis of the correlation with HAMD scores yielded Spearman coefficients of 0.614 and 0.616 for the PHQ-9 and PHQ-8, respectively, which were both favorable results.

ROC analysis of the PHQ-8 and PHQ-9

In the ROC analysis of the PHQ-8 and PHQ-9, the AUCs were 0.76 (95% CI=0.72–0.80) and 0.76 (95% CI=0.73–0.80), respectively. With a cutoff score of 10 points, the PHQ-8 showed a sensitivity and specificity of 58% and 83%, respectively; likewise, those of the PHQ-9 were 56% and 88%. The PHQ-8 showed a positive predictive value (PPV) of 53% and a negative predictive value (NPV) of 86%, and the PHQ-9 showed a PPV of 53% and an NPV of 89%. The positive and negative likelihood ratios for the PHQ-8 were 1.99 and 0.29, respectively; similarly, those for the PHQ-9 were 1.98 and 0.22. A detailed description is provided in Table 3. In the comparison of ROC curves, there was no statistically significant difference in the AUCs between the PHQ-8 and PHQ-9 (Figure 1).

DISCUSSION

The main finding of this study was that the ability of the PHQ-8 to diagnose and assess MDD is comparable to that of

the PHQ-9. There was no difference in the operating characteristics of the PHQ-8 and PHQ-9 for discriminating MDD. There is no harm in using the PHQ-8 in place of the PHQ-9 to screen for MDD.

The PHQ-8 showed excellent validity and reliability for diagnosing MDD in our sample. Indeed, the PHQ-8 showed negligible difference with the validity and reliability of the PHQ-9 in the same sample. The PHQ-9 has previously been validated in Korean studies in primary care and non-clinical settings, such as the general elderly population and students.^{19,20,25} Our study confirmed that the PHQ-8 and PHQ-9 are both valid in psychiatric clinical settings in Korea.

Our study is the first to suggest a cutoff point for the PHQ-8 in a Korean sample. Previously, in Korean samples, a cutoff point for MDD had only been suggested for the PHQ-9, and that too only in few studies. These cutoffs were 9 points for a small sample in a primary care setting, and 5 points in a study of the elderly.^{19,20} In our study, the cutoff point for both the PHQ-8 and PHQ-9 was 10. It is encouraging that the PHQ-8 and PHQ-9 also showed the same optimal cutoff points, and this should be useful because it means that MDD can be screened by using the same score on the PHQ-8 and PHQ-9. At the very least, our study presents the possibility of using a cutoff of 10 points for the PHQ-8 or PHQ-9 in the diagnosis of MDD in psychiatric outpatient units.

In our study, the PHQ-8 showed reasonable sensitivity and NPV, indicating that it is useful for screening MDD in a psychiatric outpatient setting. However, the specificity and PPV

Table 2. Convergent validity of the PHQ-8 and PHQ-9

| Tool | PHQ-9 | PHQ-8 | HAMD |
|-------|-------|-------|------|
| PHQ-9 | 1 | | |
| PHQ-8 | 0.984 | 1 | |
| HAMD | 0.614 | 0.616 | 1 |

Presented numbers are Spearman's coefficients. All correlations are significant at $p < 0.001$. PHQ: Patient Health Questionnaire, HAMD: Hamilton Depression Rating Scale

Table 3. Screening ability on the cut-off point 10 of the PHQ-8 and PHQ-9 for major depressive disorder

| | PHQ-8 | PHQ-9 |
|--------------|------------------|------------------|
| Cutoff point | ≥10 | ≥10 |
| AUC | 0.76 (0.72–0.80) | 0.76 (0.73–0.80) |
| Sensitivity | 0.83 (0.77–0.88) | 0.88 (0.83–0.92) |
| Specificity | 0.58 (0.53–0.64) | 0.56 (0.50–0.61) |
| PPV | 0.53 | 0.53 |
| NPV | 0.86 | 0.89 |
| LR+ | 1.99 | 1.98 |
| LR- | 0.29 | 0.22 |

95% confidence interval. AUC: area under the curve, PPV: positive predictive value, NPV: negative predictive value, LR+: likelihood ratio positive, LR-: likelihood ratio negative, PHQ: Patient Health Questionnaire

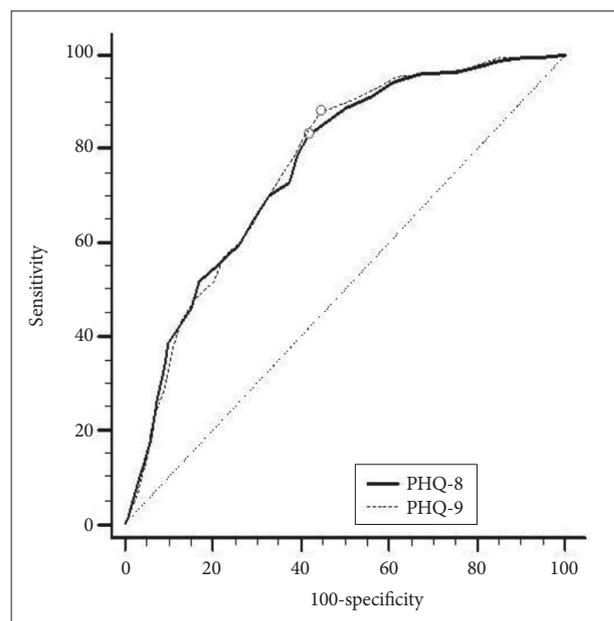


Figure 1. The receiver operating characteristic curve of the PHQ-8 and PHQ-9 versus the Mini-International Neuropsychiatric Interview for major depression diagnosis. PHQ: Patient Health Questionnaire.

were a little low. The negative outcomes that can arise because of low specificity are that a person without MDD could be found to be at risk or actually depressed in screening. This is thought to be because our sample was extracted from psychiatric outpatients, similar to previous research in a specialized psychiatric unit with a high prevalence of mood disorders.²⁶ Likewise, in our study, it is thought that patients diagnosed with adjustment disorder, unspecified depressive disorder, or panic disorder could have been false-positive cases. At a cutoff point of 10, the PHQ-8 had a slightly lower NPV compared to the PHQ-9, which means that patients with a depression test (PHQ-9 score > 10) are more likely to be depression-free when tested with the PHQ-9. In this case, lowering the cutoff point in the PHQ-8 can lead to an increase in the NPV, but a lower cutoff point inevitably leads to a lower PPV,²⁷ and a cutoff point of 10 was considered reasonable in this study.

Convergent validity with the HAMD was found to be 0.614 and 0.616 for the PHQ-8 and PHQ-9, respectively, which is considered as moderate convergent validity. Convergent validity is the degree to which two measures of constructs that theoretically should be related, are in fact related. Because convergent validity is examined by using different measures, perfect correlation cannot be expected.²⁸ There is no gold standard for the degree of convergent validity, but according to the literature, a correlation coefficient of 0.6 or greater can be considered as strong evidence supporting construct validity.²⁸ In addition, the correlation coefficient between the instruments used in our study sample was similar to previous studies showing a convergent validity coefficient of 0.52–0.56 between the HAMD and PHQ-9.^{29,30} Thus, the PHQ-8 and PHQ-9, similar to the HAMD, can be said to have a significant correlation in their ability to measure depression.

The ability to screen for MDD is important in national health examination, primary care screening, and cohort studies. The PHQ-9 was determined to be appropriate for screening for MDD, but Item 9 was only ambiguous to actually assess suicidal ideation. However, it is very difficult to provide immediate psychiatric services in the abovementioned settings to individuals who respond with more than 1 point to item 9. Therefore, it may be appropriate to use the PHQ-8 in MDD screening if the ability of Item 9 to assess suicidal ideation is unclear.

Our study had the following limitations. As the sample was extracted from a psychiatric outpatient clinic, it may be difficult to extend our results to the general population. If the PHQ-8 was validated in the general population in later studies, it would provide better evidence for screening in the general population. In total, 59.3% of the subjects in this study responded with a score of 1 point or higher for Item 9. This is much higher than the rate of 7.3% among the general Korean

population in a previous study we conducted. The current prevalence of suicidal ideation is thought to be higher because the subjects were visiting a psychiatric clinic. If research is conducted in the general population, we expect that the PHQ-8, which excludes Item 9, will demonstrate even better ability to assess and screen for MDD.

In conclusion, the PHQ-8 shows sufficient validity to be used in screening for MDD. In addition, it shows almost no difference in its ability to screen for MDD when compared to the PHQ-9.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HC15C1405).

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Han CS, Shin CM. Data curation: Shin CM, Lee SH. Formal analysis: Shin CM. Investigation: Shin CM, Yoon HK, Han KM. Writing—original draft: Shin CM. Writing—review & drafting: Han CS, Yoon HK.

ORCID iDs

Changsu Han <https://orcid.org/0000-0002-4021-8907>
Cheolmin Shin <https://orcid.org/0000-0002-8232-2921>

REFERENCES

- Gili M, Garcia Toro M, Armengol S, Garcia-Campayo J, Castro A, Roca M. Functional impairment in patients with major depressive disorder and comorbid anxiety disorder. *Can J Psychiatry* 2013;58:679-686.
- Angst J, Angst F, Stassen HH. Suicide risk in patients with major depressive disorder. *J Clin Psychiatry* 1999;60(Suppl 2):57-62; discussion 75-56, 113-116.
- Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. *Am J Orthopsychiatry* 2007;77:534-542.
- Shim RS, Baltrus P, Ye J, Rust G. Prevalence, treatment, and control of depressive symptoms in the United States: results from the National Health and Nutrition Examination Survey (NHANES), 2005-2008. *J Am Board Fam Med* 2011;24:33-38.
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606-613.
- Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med* 2010;8:348-353.
- van Steenbergen-Weijenburg KM, de Vroeghe L, Ploeger RR, Brals JW, Vloedveld MG, Veneman TF, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. *BMC Health Serv Res* 2010;10:235.
- Strine TW, Mokdad AH, Balluz LS, Gonzalez O, Crider R, Berry JT, et al. Depression and anxiety in the United States: findings from the 2006 Behavioral Risk Factor Surveillance System. *Psychiatr Serv* 2008;59:1383-1390.
- Shin C, Kim Y, Park S, Yoon S, Ko YH, Kim YK, et al. Prevalence and associated factors of depression in general population of Korea: results from the Korea National Health and Nutrition Examination Survey,

2014. *J Korean Med Sci* 2017;32:1861-1869.
10. Walker J, Hansen CH, Butcher I, Sharma N, Wall L, Murray G, et al. Thoughts of death and suicide reported by cancer patients who endorsed the "suicidal thoughts" item of the PHQ-9 during routine screening for depression. *Psychosomatics* 2011;52:424-427.
 11. Razykov I, Ziegelstein RC, Whooley MA, Thombs BD. The PHQ-9 versus the PHQ-8--is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study. *J Psychosom Res* 2012;73:163-168.
 12. Pressler SJ, Subramanian U, Perkins SM, Gradus-Pizlo I, Kareken D, Kim J, et al. Measuring depressive symptoms in heart failure: validity and reliability of the patient health questionnaire-8. *Am J Crit Care* 2011;20:146-152.
 13. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord* 2009;114:163-173.
 14. McGuire LC, Strine TW, Allen RS, Anderson LA, Mokdad AH. The Patient Health Questionnaire 8: current depressive symptoms among U.S. older adults, 2006 Behavioral Risk Factor Surveillance System. *Am J Geriatr Psychiatry* 2009;17:324-334.
 15. Wells TS, Horton JL, LeardMann CA, Jacobson IG, Boyko EJ. A comparison of the PRIME-MD PHQ-9 and PHQ-8 in a large military prospective study, the Millennium Cohort Study. *J Affect Disord* 2013;148:77-83.
 16. Bagalkot TR, Park JI, Kim HT, Kim HM, Kim MS, Yoon MS, et al. Lifetime prevalence of and risk factors for suicidal ideation and suicide attempts in a Korean community sample. *Psychiatry* 2014;77:360-373.
 17. Kim SW, Yoon JS. Suicide, an urgent health issue in Korea. *J Korean Med Sci* 2013;28:345-347.
 18. 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.
 19. 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.
 20. Choi HS, Choi JH, Park KH, Joo KJ, Ga H, Ko HJ, et al. Standardization of the Korean Version of Patient Health Questionnaire-9 as a Screening Instrument for Major Depressive Disorder. *J Korean Acad Fam Med* 2007;28:114-119.
 21. Smith MV, Gotman N, Lin H, Yonkers KA. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? *Gen Hosp Psychiatry* 2010;32:544-548.
 22. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
 23. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278-296.
 24. 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.
 25. Yoon S, Lee Y, Han C, Pae CU, Yoon HK, Patkar AA, et al. Usefulness of the patient health questionnaire-9 for Korean medical students. *Acad Psychiatry* 2014;38:661-667.
 26. Inoue T, Tanaka T, Nakagawa S, Nakato Y, Kameyama R, Boku S, et al. Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. *BMC Psychiatry* 2012;12:73.
 27. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilitys, and pitfalls in research and practice. *Front Public Health* 2017;5:307.
 28. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials: a review. *Health Technol Assess* 1998;2(14).
 29. Wittkamp K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009;31:451-459.
 30. Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. *BMC Psychiatry* 2008;8:46.