The Psychometric Properties of the Persian Versions of the Patient Health Questionnaires 9 and 2 as Screening Tools for Detecting Depression among University Students

Abstract

Background: Among the common mental disorders in societies, depression is one of the most common mental disorders that affects all groups and classes of society. Students are among the groups with the highest rates of depression. Therefore, the need for a short and effective tool for screening and early detection of depression is felt. The aim of this research is to determine validity, reliability and the best cut-off point of the patient health questionnaires-9 (PHQ-9) and patient health questionnaires-2 (PHQ-2) in university students. **Methods:** This cross-sectional study was conducted on 246 students of Kermanshah University of medical science in Kermanshah province of Iran. They completed the PHQ-2, PHQ-9, and the Beck Depression Inventory-II (BDI-II). A structured interview was used to diagnose depression. To analyze the data, Cronbach's alpha for internal consistency, the intra-class correlation (ICC) for test–retest reliability, confirmatory factor analysis for construct validity, Pearson Correlation for Convergent validity, and receiver-operating characteristic (ROC) curve for Criterion validity was used. **Results:** The mean age of the participants was 20.43 ± 2.29. Cronbach's alpha coefficient for PHQ-9 and PHQ-2 was 0.82 and 0.80, respectively. The test–retest reliability based on intra-class correlation (ICC) for PHQ-9 and PHQ-2 after two weeks was 0.81 and 0.73, respectively ($P < 0.001$). The correlation coefficient between the PHQ-9 and PHQ-2 with the BDI-II was 0.74 and 0.64, respectively ($P < 0.001$). Confirmatory factor analysis showed that two-factor model and one factor model had good model fit. The best cut-off point score for the PHQ-9 was 10 with a sensitivity of 0.90 and specificity of 0.93, and the best cut-off point score for the PHQ-2 was 3 with the sensitivity of 0.71 and specificity of 0.92. **Conclusions:** The PHQ-9 and PHQ-2 are suitable tools to screen depression in the university students in Iran.

Keywords: *Depression, patient health questionnaires, psychometric, University students*

Rasoul Mohamadian, Habibolah Khazaie[1], Seyed Mojtaba Ahmadi[2], Mehdi Fatmizade[3], Susan Ghahremani[4], Hossein Sadeghi[5], Zohreh Ansari[4], Ahmadreza Ahmadi[6] Sadighe Zare[2]

Department of Clinical Psychology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, [1]Sleep Disorders Research Center, Kermanshah University of Medical Sciences (KUMS), Kermanshah, [2]Department of Clinical Psychology, School of Medicine, Kermanshah University of Medical Sciences (KUMS), Kermanshah, [3]Department of Clinical Psychology, School of Medicine, Tehran University of Medical Sciences, Tehran, [4]Department of Clinical Psychology, School of Medicine, Iran University of Medical Sciences, Tehran, [5]Department of Clinical Psychology, School of Medicine, Ghazvin University of Medical Sciences, Ghazvin, [6]Medical Department, School of Medicine, Jondishahpour University of Medical Sciences, Ahvaz, Iran

Address for correspondence:
Dr. Seyed Mojtaba Ahmadi, Department of Clinical Psychology, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran.
E-mail: mojtabakmahmadi@ yahoo.com

Introduction

Depression is correlated with a high level of mortality, sickness, and disability.[1] The prevalence of the lifetime of the depressive disorder is from 5% to 17%. The mean age of the onset of Major Depressive Disorder (MDD) is approximately 40 and about 50% of people experience the onset of MDD between the ages of 20 and 50.[2] According to studies, university students are among the groups that have been observed to have a high level of depression. Based on a systematic review and meta analysis,the prevalence of depression or depressive symptoms among medical students was27.2%.[3] Depression can cause loss of motivation, disruption in daily activities, suicide, drug abuse, academic failure, etc.[4-6] Given the prevalence of

depression among students and since students in each country will be the future makers of society, special attention should be paid to this issue and emphasis should be placed on its rapid diagnosis and treatment. Therefore, fast and correct diagnosis of depression is an important necessity in societies as it can play a preventative role and be economically advantageous. The availability of a reliable and easy-to-implement screening tool is an important factor to the success of screening for depression.[7] Different tools such as the Zung Self-Rating Depression Scale (SDS), the depression subscale of the SCL-90, the depression subscale of the Hopkins Symptom Checklist (HSCL), and the Beck Depression Inventory have been utilized to assess depression in medical college students.[8] The right screening tool should

have characteristics such as briefness, easy admiration, and culturally acceptable to the to the subjects.[7] The PHQ-9 questionnaire is a self-report questionnaire that consists of 9 items which are based on DSM-IV criteria and it takes half as long as the other questionnaires that were already mentioned.[9,10] Therefore, it has priority over other methods of screening. The PHQ-2 contains the first 2 questions of the PHQ-9, and as a result, it can be used as a brief depression screening tool and this way the responsibility of screening depression is greatly minimized.[10]

The PHQ-9 has been translated in different countries such as China,[7] Lebanon(Arabic speaking Lebanese),[11] Japan[12-13] and has shown an very well reliability. The studies have shown that the optimal cut-off score for the PHQ-9 might vary for different samples and targets.[11,14] Although the studies have proven the reliability of the PHQ-9 as a screening tool, this questionnaire has not been evaluated for psychometric properties in Iranian college students. As a result, this study is aiming at evaluating the reliability and validity of PHQ-9 and PHQ-2 as screening tools for university students in Iran.

Materials (subjects) and Methods

Participants

The participants in the present study were from the Kermanshah University of Medical Science, Kermanshah, Iran. The number of participants was 246 students who were selected using a convenience sampling method. According to Myers *et al.*,[15] The sample size of 200 people is good for confirmatory factor analysis, so considering the possibility of a drop-out, We considered the sample size to be 300 students, but 54 students were excluded from the study due to incomplete questionnaires or incomplete clinical interviews. After the introduction of the study and obtaining consent, the questionnaires PHQ-9, PHQ-2, and BDI-II were distributed to the students and were later collected.

Procedures

Four of the co-researchers who either held master's degrees or Ph.D.'s and had received proper training in the area of psychiatric interviews conducted the interview. The research assistants used DSM-V in administering the Structured Clinical Interview (SCID). The interviews were conducted on a face-to-face basis in the clinic of the Farabi Hospital and other university counseling centers of Kermanshah University of Medical Science in Kermanshah. Each interview lasted from 30 to 45 minutes on average. After the interview, the students who were diagnosed with depression or had suicidal thoughts were referred to the university counseling centers. Next, the students who provided their consent and could participate in the research completed the questionnaires PHQ-9, PHQ-2, and BDI-II.

Measures

The patient health questionnaire-9 (PHQ-9) and the patient health questionnaire-2 (PHQ-2): The PHQ-9 is a self-report scale that is developed to make a diagnosis based on the DSM-4. This questionnaire has 9 questions and has originated from the full PHQ. The participants must rate what they have been feeling in the last two weeks on a Likert 4 point scale (never: 0, some days: 1, more than half the days: 2, almost every day: 3). The range of the scores is between 0 and 27; a higher score indicates that the signs of depression are more severe. The PHQ-2 consists of the first two questions of the PHQ-9 that analyze low mood and lack of interest in enjoyable activities in the last two weeks; each question is given a weight of 0 to 3 (never: 0, some days: 1, more than half the days: 2, almost every day: 3) and the range of the scores is from 0 to 6.[9,10]

The Structured Clinical Interview (SCID): The Structured Clinical Interview for the DMS-V (SCID-5) is a semi-structured interview for making the major DMS-5 diagnoses, which was put in Axis-1 in the past. This tool needs to be administered by a trained clinician or mental health professional that is familiar with the diagnosis criteria and the classification of the disorders in DMS-5.[16] SCID was used to a gold standard to diagnose depression.

The Beck Depression Inventory (BDI-II): This questionnaire is a 21-item self-report tool for measuring the severity of depression.The scoring system is based on a Likert scale that is assigned a score from 0 to 3. A higher score indicates that depression is more severe.[17] This questionnaire has been utilized in Iran and has shown an acceptable level of reliability. Namely, In a study was conducted in Iran Cronbach's alpha was 0.87.[18]

Statistical analysis

The statistical software packages SPSS Statistics 21, Lisrel 8.8, and Medcalc were used to analyze the data. In the analysis of the reliability data, Cronbach's alpha method was utilized to examine the internal consistency. Alpha coefficient greater than 0.7 is in the acceptable range.[19] The intra-class correlation (ICC) was used for test–retest reliability. ICC <0.5 indicates poor reliability, 0.50 to 0.75 indicates moderate reliability, 75-90 indicates good reliability and higher than 0.90 indicates excellent reliability.[20]

For the part of validity, to analyze Convergent validity, BDI-II correlation with PHQ-9 and PHQ-2 was utilized and the Pearson correlation coefficient was calculated for this purpose. To interpret the correlation coefficient, correlation values of 0.40 or above were considered satisfactory ($r \geq 0.81$-1.0 as excellent, 0.61-0.80 very good, 0.41–0.60 good, 0.21–0.40 fair, and 0–0.20 poor).[21]

Confirmatory factor analysis was used to examine construct validity. The Lisrel software was used for confirmatory factor analysis. Indicators reported for model fit are the

goodness of fit index (GFI), the root mean square error of approximation (RMSEA), comparative fit index (CFI), and χ^2/df. Because the Chi-square statistic is sensitive to the sample size, the Chi-square statistic along with the degree of freedom (X^2/df) was calculated for the overall evaluation of the model. A value of X^2/df ≤ 2 indicates a good model fit and $2 < X^2$/df ≤ 3 indicates an acceptable model fit. RAMSEA between 0 and 0.05 indicate a good fit, and 0.05 to 0.08 indicate an acceptable fit. For GFI, 0.90 and above indicate acceptable fit and 0.95, and above indicates good fit. For CFI, $0.95 \leq$ CFI < 0.97 indicates an acceptable model fit and $0.97 \leq$ CFI ≤ 1.0 indicate good fit.[22]

Criterion validity of PHQ-9 and PHQ-2 was evaluated using the receiver-operating characteristic (ROC) curves in this study. The following categories were used to interpret AUC: 90 -100 excellent; 80–90 good; 70–80 fair; 60–70 poor; 50–60 fail.[23] To determine the optimal cut-off point, the sensitivity, specificity, and area under the ROC curve (AUC) were calculated. For this purpose, the clinical interview with SCID was considered as the gold standard for diagnosing depression. Therefore, we had two groups with a diagnosis of depression and without a diagnosis of depression. In order to obtain the optimal cut-off point, the maximal Youden Index (J = sensitivity + specificity – 1) was used.[24] All tests were two-tailed, and $P < 0.05$ was considered significant.

Ethical considerations

This research project was approved by the Ethics Committee of Kermanshah University of Medical Sciences.

Results

Properties of the sample

In total, 54 out of 300 students who were screened during this study did not return for the clinical interview (SCID). Hence, 246 of the participants i.e., 82% of them completed the process of the study. Of the participants, 122 (49.6%) were men and 124 (50.4%) were women. The mean age of the participants was 20.43 ± 2.29 in the range of 18 to 33. 17 (6.9%) of the participants were studying clinical psychology, 135 (54.9%) were studying nursing, 62 (25.2%)

were studying anesthesiology, 21 (8.53%) were studying operating room technology and 11 (4.47%) were studying public health. Of this number, 207 (84.14%) were completing the undergraduate studies, 29 (11.8%) were completing the master's degree and 10 (4.06%) were Ph.D. students.

Reliability analysis

The internal consistency of the PHQ-9 based on Cronbach's alpha was 0.82 and the item-total correlation between each item and the total score was determined to be in the range of 0.5 to 0.73 ($P < 0.001$). The internal consistency of the PHQ-2 based on Cronbach's alpha was 0.80 and the item-total correlation between each item and the total score was determined to be in the range of 0.89 to 0.90 ($P < 0.001$). Based on the analysis of the items of the questionnaire, it was found that the omission of every question reduces the value of alpha, which is an indication of the questions' suitability, except item number 3, the omission of which would not change the value of alpha [Table 1].

The test–retest reliability based on intra-class correlation (ICC) was 0.81 (ICC = 0.81,95%CI = 0.73-0.87, $P < 0.001$) and 0.73 (ICC = 0.73, 95%CI = 0.60-0.82, $P < 0.001$), respectively.

Validity

Convergent validity

The correlation coefficient between the PHQ-9 and PHQ-2 with BDI-II was 0.74 and 0.64, respectively ($P < 0.001$).

Construct validity

Confirmatory factor analysis was used to examine construct validity. The Lisrel software was used in order to do confirmatory factor analysis. In some of the previous studies, the one-factor and two-factor models (cognitive and somatic) were suggested for this questionnaire.[25] Both models were examined in this article. The results indicate that that the two-factor model (CFI: 1, GFI = 98, RAMSEA = 0.017, $\chi 2$ = 27.76, df = 26) had a better fit than one-factor model (CFI: 0.97, GFI = 0.95, RAMSEA = 0.074, $\chi 2$ = 63.71, df = 27). The results of the examination have shown that both models have good fitness [Figure 1].

item		Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
p1	Little interest or pleasure in doing things	0.67	0.78
p2	Feeling down, depressed, or hopeless	0.69	0.77
p3	Trouble falling or staying asleep, or sleeping too much	0.37	0.82
p4	Feeling tired or having little energy	0.45	0.81
p5	Poor appetite or overeating	0.47	0.81
p6	Feeling bad about yourself—or that you are a failure	0.60	0.79
p7	Trouble concentrating on things	0.51	0.80
p8	Moving or speaking so slowly that other people could have noticed	0.42	0.81
p9	Thoughts that you would be better off dead or of hurting yourself	0.51	0.80

Table 1. Item analysis of PHQ-9

Chi-Square=63.71, df=27, P-value=0.00008, RMSEA=0.074

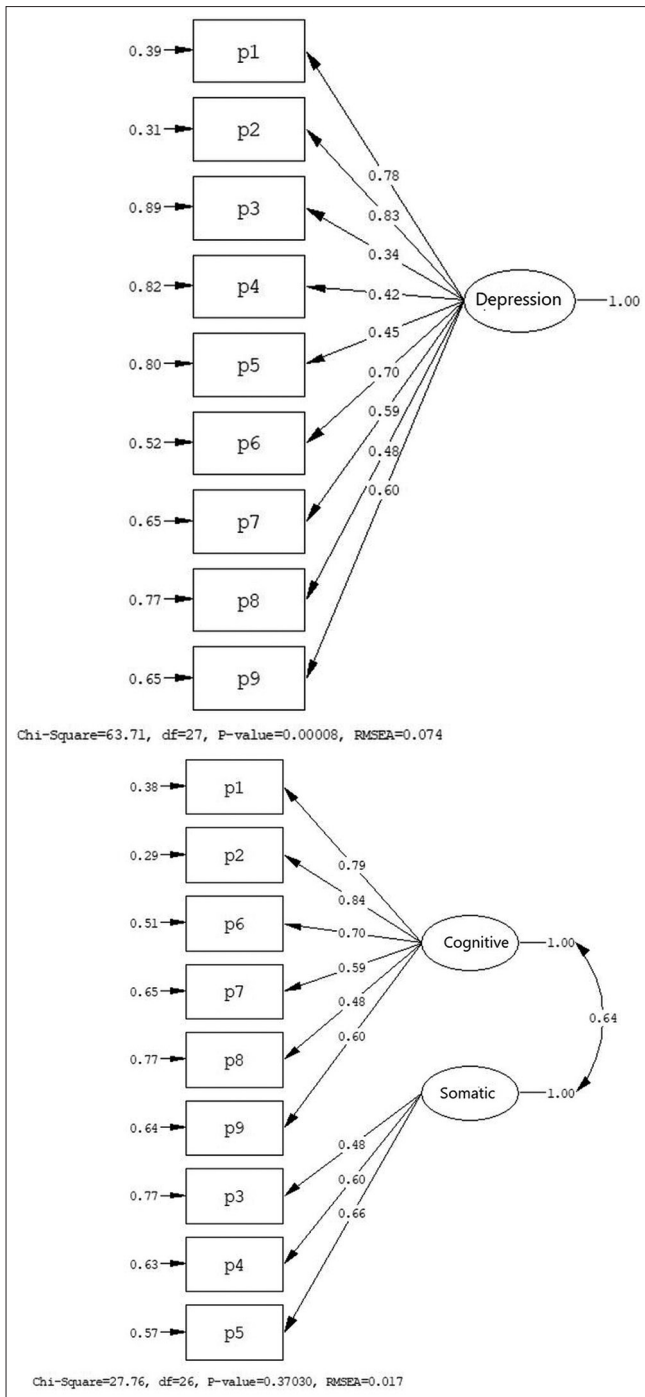Chi-Square=27.76, df=26, P-value=0.37030, RMSEA=0.017

Figure 1: One-factor and two-factor models of the PHQ-9 items, P1: Little interest or pleasure in doing things, p2: Feeling down, depressed, or hopeless; P3: Trouble falling or staying asleep, or sleeping too much; P4: Feeling tired or having little energy, P5: Poor appetite or overeating, P6: Feeling bad about yourself—or that you are a failure; P7: Trouble concentrating on things; P8: Moving or speaking so slowly that other people could have noticed P9: Thoughts that you would be better off dead or of hurting yourself

Criterion validity

Characteristics of a screening test

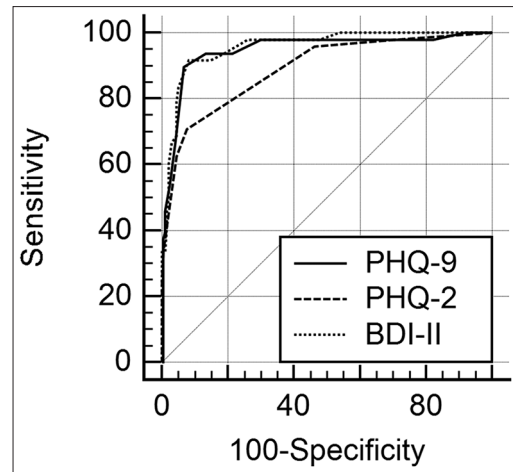The ROC test was used to obtain the best cut-off point. The results of this test indicate that The AUC for PHQ-9

Figure 2: Receiver-operating characteristic curves for PHQ-9, PHQ-2 and BDI-II (n = 246)

Table 2: Area under the curves (AUC) for PHQ-9, PHQ-2 and BDI-II

Variable	AUC	Standard error (SE)	95% confidence interval
PHQ-9	0.947	0.0206	0.910 to 0.971
PHQ-2	0.890	0.0263	0.844 to 0.926
BDI-II	0.957	0.0143	0.924 to 0.979

was 0.947 (AUC = 0.974, 95%CI = 0.91-0.97), for PHQ-2 was 0.89 (AUC = 0.89, 95%CI = 0.84-0.92) and for BDI-II was 0.957 (AUC = 0.957, 95%CI = 0.92-0.979) [Figure 2 and Table 2]. Medcalc software was used to compare the area under the curves. The results showed that PHQ-9 and BDI-II have more AUC then PHQ-2 ($P < 0.05$). Compared to BDI-II and PHQ-9, although the area under the curve was higher for BDI-II, this difference was not statistically significant ($P = 0.49$) [Table 3].

The results have shown that of the 246 participants, 48 (19.52%) were depressed and 198 (80.48%) non-depressed. The best cut-off point for the PHQ-9 among the university students in Iran was 10 with the sensitivity of 0.90 and specificity of 0.93, and the best cut-off point for the PHQ-2 among the university students in Iran was 3 with the sensitivity of 0.71 and the specificity of 0.92 [Table 4].

Discussion

The present study aimed to determine the psychometric characteristics of the PHQ as a brief screening tool among the Iranian university student population. This study was done to examine psychometric properties of the PHQ-9 and PHQ-2 among Iranian university students. The results of our study showed that PHQ-9 and PHQ-2 questionnaires have good reliability and validity. Although both one-factor and two-factor models had a good fit, the two-factor model had a better fit. The best cut-off points for PHQ-9 and PHQ-2 questionnaires were 10 and 3, respectively. PHQ-9 and BDI-II have more AUC (Accuracy) than PHQ-2. Compared to BDI-II and PHQ-9, although the area under

Table 3: Pairwise comparison of ROC curves

	PHQ-9 ~PHQ-2	PHQ-9 ~BDI-II	PHQ-2 ~BDI-II
Difference between areas	0.0563	0.0109	0.0672
Standard Error	0.0250	0.0160	0.0244
95% Confidence Interval	0.00731-0.105	-0.0205-0.0423	0.0194-0.115
Z statistic	2.252	0.680	2.754
Significance level	*P*=0.0243	*P*=0.4967	*P*=0.0059

Table 4: Operating characteristics of various cut-off scores of the PHQ-9 for diagnosing major depression among university student

Test result variable (s)	Positive if Greater than or Equal Toa	Sensitivity	Specificity	PPV*	NPV*	LR+*	LR-*
PHQ-9	≥9	0.94	0.86	0.63	0.98	7.10	0.07
	≥10	0.90	0.93	076	0.97	13.57	0.11
	≥11	0.67	0.95	0.8	0.92	16.26	0.34
PHQ-2	≥2	0.96	0.53	0.36	0.98	2.08	0.07
	≥3	0.71	0.92	0.69	0.92	9.31	0.31

*PPV=Positive Predictive Value; NPV=Negative Predictive Value; LR+=Positive likelihood ratio; LR-=Negative likelihood ratio

the curve was higher for BDI-II, this difference was not statistically significant.

The reliability of the PHQ-9 and PHQ-2 based on Cronbach's alpha has been determined to be 0.82 and 0.80, respectively. This is in line with the two studies that were done on the students in Chinas (Cronbach's alpha for PHQ-9 and PHQ-2 = 0.85, 0.72, respectively) and Koreans (Cronbach's alpha for PHQ-9 = 0.837).[7,26] The present study is also in line with the study done on the general public. In the above-mentioned study, the value of Cronbach's alpha for the PHQ-9 had measured 0.87.[27] The test–retest reliabilities (based on intra-class correlation (ICC)) of both questionnaires were high and measured 0.81 and 0.73, respectively. This is consistent with the studies done on students in the Chinas (The intra-class correlation coefficient for the test–retest reliability of PHQ-9 and PHQ-2 were 0.87 and 0.82, respectively).[7] An ICC of 75-90 reflects good reliability.[20] Therefore, the reported reliability coefficient indicates the suitability of this tool for screening purposes among university students.

The high correlation between PHQ-9, PHQ-2, and BDI-II is indicative of their concurrent validity for screening purposes and this in itself is a confirmation of the previous studies. Similar results have been achieved in the study that was done on the student sample in China. In the aforementioned study, the correlation coefficient between the PHQ-9 and BDI-II measured 0.79 and the correlation coefficient between the PHQ-2 and BDI-II measured 0.65.[7]

Both the single-factor and two-factor models have been obtained in this study; however, the two-factor model has better fitness. In some of the previous studies, a single-factor model has been obtained.[7] However, in some other studies the two-factor model has been acquired.[11,28] And in some studies both the single-factor and the two-factor models have been suggested.[25]

The score of the optimal cut-off point for the PHQ-9 is equal to 10 (sensitivity = 0.90 and specificity = 0.93). AUC of PHQ-9, which were excellent according to the conventional classification system.[23] Based on the obtained AUCs, the Persian versions of PHQ-9 had the discriminatory ability to correctly classify students with and without depression.

In a meta-analysis that has assessed the optimal cut-off point for the PHQ-9, 18 studies, in this area, have been examined and the result was indicative of the fact that the optimal cut-off point could be in the range of 8 to 11. In this range, the PHQ-9 can be used as a depression diagnostic.[29] This shows that the cut-off score will vary based on the target population.[11]

The optimal cut-off point for the PHQ-2 is equal to 3. This cut-off score has a sensitivity of 0.71 and a specificity of 0.92. This is indicative of the adequacy of PHQ-2 as a quick screening tool. The acquired results are aligned with the studies that were done in China. In the Chinese study, the optimal cut-off score for university students was determined to be 3 with a sensitivity of 0.81 and specificity of 0.96.[7]

Limitations

Sample size and convenience sampling method were among the limitations of this study. This study was performed on a sample of 246 students from Kermanshah University of Medical Sciences who were selected by convenience sampling method.

We have some limitations in generalizing the results to other communities. It is suggested that other studies be conducted on other communities with a larger sample size and, if possible, using random sampling methods.

Conclusion

The results from the present study indicated that PHQ-9 and PHQ-2 questionnaires have suitable reliability and validity and accuracy among the Iranian university students. The PHQ-9 and PHQ-2 with cut-off points of 10 and 3, respectively, can be used for diagnostic and screening purposes.

Conflicts of interest

There are no conflicts of interest.

References

1. Weinberger A, Gbedemah M, Martinez A, Nash D, Galea S, Goodwin R. Trends in depression prevalence in the USA from 2005 to 2015: Widening disparities in vulnerable groups. Psychol Med 2018;48:1308-15.

2. Sadock BJ, Sadock VA, Ruiz P. Kaplan and Sadock's synopsis of psychiatry: Behavioral sciences/clinical psychiatry, 11th ed. Wolters Kluwer Health; 2015.

3. Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, *et al*. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: A systematic review and meta-analysis. JAMA 2016;316:2214-36.

4. Loloie M, Haghdoost AA, Bazrafshan A, Okhovati M, Mokhtari AM, Zare M. Factors affecting undergraduate students' academic failure in Kerman University of Medical Sciences, Iran; A case-control study. Strides in Development of Medical Education 2017;13:505-14.

5. Daraei M. An investigation of the relationship between depression and self-esteem with academic performance. Yafteh 2018;19:117-26.

6. Azarniveh M, Tavakoli Khormizi S. Depression and academic achievements in athletes and non-athletes university students: A comparative study. Iran J Nurs 2015;28:1-10.

7. Zhang YL, Liang W, Chen ZM, Zhang HM, Zhang JH, Weng XQ, *et al*. Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China. Asia Pac Psychiatry 2013;5:268-75.

8. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. Acad Med 2006;81:354-73.

9. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16:606-13.

10. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: Validity of a two-item depression screener. Med Care 2003;41:1284-92.

11. Sawaya H, Atoui M, Hamadeh A, Zeinoun P, Nahas Z. Adaptation and initial validation of the Patient Health Questionnaire–9 (PHQ-9) and the Generalized Anxiety Disorder–7 Questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample. Psychiatry Res 2016;239:245-52.

12. Umegaki Y, Todo N. Psychometric properties of the Japanese CES–D, SDS, and PHQ–9 depression scales in university students. Psychol Assess 2017;29:354-9.

13. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Tanaka Y, Hosaka M, *et al*. Performance of the Japanese version of the Patient Health Questionnaire-9 (J-PHQ-9) for depression in primary care. Gen Hosp Psychiatry 2018;52:64-9.

14. Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. Gen Hosp Psychiatry 2010;32:345-59.

15. Myers ND, Ahn S, Jin Y. Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. Res Q Exerc Sport 2011;82:412-23.

16. First MB, Williams J, Karg RS, Spitzer RL. User's Guide to Structured Clinical Interview for DSM-5 Disorders (SCID-5-CV) Clinical Version. Arlington, VA: American Psychiatric Publishing; 2015.

17. Beck AT, Steer RA, Brown GK. Beck depression inventory-II. San Antonio 1996;78:490-8.

18. Rahimi C. Application of the beck depression inventory-II in Iranian university students. Clinical Psychology & Personality 2014;2:173-88.

19. Leech NL, Barrett KC, Morgan GA. SPSS for Intermediate Statistics: Use and Interpretation. Psychology Press; 2005.

20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine. 2016;15:155-63.

21. Dancey CP, Reidy J. Statistics without maths for psychology, 7th ed. Pearson education; 2017.

22. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research Online 2003;8:23-74.

23. Safari S, Baratloo A, Elfil M, Negida A. Evidence based emergency medicine; part 5 receiver operating curve and area under the curve. Emergency 2016;4:111.

24. Stafford L, Berk M, Jackson HJ. Validity of the hospital anxiety and depression scale and patient health questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry 2007;29:417-24.

25. González-Blanch C, Medrano LA, Muñoz-Navarro R, Ruíz-Rodríguez P, Moriana JA, Limonero JT, *et al*. Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. PLoS one 2018;13:e0193356.

26. 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;38:661-7.

27. Kocalevent R-D, Hinz A, Brähler E. Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2013;35:551-5.

28. Granillo MT. Structure and function of the Patient Health Questionnaire-9 among Latina and non-Latina White female college students. Journal of the society for social work and Research 2012;3:80-93.

29. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. CMAJ 2012;184:E191-6.