

Short Communication

The Reliability and Validity of the Hebrew Patient Health Questionnaire (PHQ-9) in the General Population

Tomer Yona¹, Asaf Weisman², Uri Gottlieb³, Eshed Lin², Youssef Masharawi²

¹ Department of Biomedical Engineering, Technion, Israel Institute of Technology, Haifa, Israel

² Spinal Research Laboratory, Department of Physical Therapy, Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, Israel.

³ Neuromuscular and Human Performance Laboratory, Department of Physiotherapy, Ariel University, Israel

Corresponding author

Tomer Yona

Technion, Israel Institute of Technology, Haifa, Israel

E-Mail: tomeryona@campus.technion.ac.il

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Short Communication

ABSTRACT

Objective: To assess the psychometric properties of the Hebrew version of the Patient Health Questionnaire (PHQ-9) in the general population.

Methods: Using an online survey, we assessed test-retest reliability with a two-week time interval. A total of 118 participants enrolled in the study, of whom 103 completed the survey twice. Each participant filled out the PHQ-9 and the 12-Item Short Form Survey (SF-12). Our statistical analysis includes Cronbach's alpha, Intraclass Correlation Coefficient (ICC_{2,1}), Spearman's rank correlation coefficient, Standard Error of Measurement (SEM), and Minimal Detectable Change (MDC).

Results: Internal consistency of the Hebrew version of the PHQ-9 ranged from $\alpha=0.79-0.83$. The test-retest reliability of the questionnaire is good (ICC_{2,1}= 0.81), and it is moderately and negatively correlated to the mental component of the SF-12 (Spearman $\rho= -0.57$, $p< .05$). The SEM of the PHQ-9 is 1.83 points, and the MDC was found to be 5 points.

Conclusion: The Hebrew version of the PHQ-9 is valid and reliable for screening self-reported depressive symptoms online in the general Hebrew-speaking population.

Keywords: Depressive symptoms, Hebrew PHQ-9, reliability and validity, SF-12, Online survey

Short Communication

INTRODUCTION

Depression is among the leading causes of disability worldwide. Additionally, the prevalence of depression and depressive symptoms has been rising in recent decades.^{1,2} Despite the availability of many self-reported tools to screen and diagnose perceived depressive symptoms, depression remains under-detected among various populations.³⁻⁵

The Patient Health Questionnaire (PHQ-9) was developed as a short questionnaire based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria to diagnose depressive disorders.⁶ It aims to screen self-reported depressive symptoms is freely available and has been translated and validated in many different languages and populations.⁷⁻¹⁰ However, the psychometric properties of the Hebrew version of the PHQ-9 have not been investigated.

Properly validated and reliable tests are essential for clinical decisions, diagnoses, and prognoses.¹¹ Moreover, using unvalidated or unreliable instruments can affect how authors report on a study's results.¹² Consequently, this paper aims to assess the internal consistency, test-retest reliability, and construct validity of an online administration of the Hebrew version of the PHQ-9 on the general Hebrew-speaking population in Israel.

Short Communication

METHODS

Tel Aviv University's ethics committee approved this study (Number 0003314-2). The study design followed the COSMIN study design checklist for patient-reported outcome measurement instruments.¹³

Procedures

We used an online platform (www.Alchemer.com) to disseminate the survey in June 2021. Hebrew-speaking respondents were recruited from the Israeli adult general population using social media. The first page of the survey consisted of the survey description, the informed consent, and the researcher's contact details. Inclusion criteria were a minimum age of 18 years old and the ability to read and comprehend Hebrew at a native level of language proficiency.

Each respondent completed the survey twice, with a two-week interval between testing. The first time-point included only the PHQ-9 and the second time-point included the 12-Item Short Form Survey (SF-12) and the PHQ-9. Additionally, at the second time point, respondents were first asked whether they "experienced a change in their health in the previous two weeks?" Any respondent that answered "yes" was directed out of the survey and was excluded from the reliability analysis.

Patient Health Questionnaire (PHQ-9)

The PHQ-9 is a self-reported questionnaire consisting of nine items to assess the severity of depressive symptoms. It is used to determine the presence of bothersome symptoms experienced by the participant in the last two weeks using a 0-3 Likert scale. The minimum total score (0) indicates no depressive symptoms, and the highest total score (27) indicates severe depressive symptoms. We used the official Hebrew version of the PHQ-9, freely available online (<https://www.phqscreeners.com>).

Short Communication

The 12-Item Short Form Survey (SF-12)

The SF-12 is a self-reported questionnaire consisting of twelve items to assess the respondent's general, physical, and mental health. The summary score of the SF-12 consists of a Physical Composite Scale (PCS) and a Mental Health Composite Scale (MCS). A higher score indicates a better health status. To assess the validity of the PHQ-9, we used the Hebrew version of the SF-12 that has already been proven as reliable and valid.¹⁴

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows (Version 25, Armonk, NY) and Jamovi (Version 1.6, The Jamovi project). Internal consistency was calculated using Cronbach's alpha (α), while Intraclass Correlation Coefficient ($ICC_{2,1}$) was used to calculate the test-retest reliability and Spearman's rank correlation coefficient (ρ) was used for construct validity. The following scale was used to determine ICC: higher than 0.90 - excellent reliability; above 0.75 - good reliability; and 0.50-0.75 - moderate reliability.¹⁵ For Spearman's correlations, the following scale was used: 0.7-0.9= strong; 0.4-0.6= moderate; and 0.1-0.3= weak.¹⁶ For the calculation of the Standard Error of Measurement (SEM), the pooled standard deviations formula was used ($SEM = SD * \sqrt{1 - ICC}$). The formula for the Minimal Detectable Change (MDC) was $MDC = SEM * 1.96 * \sqrt{2}$ ¹⁵. Lastly, we considered a floor and ceiling effect if more than 15% of participants achieved the highest or lowest score in the PHQ-9.¹⁷

Short Communication

RESULTS

A total of 132 respondents entered the survey in the first time point, of whom 118 respondents completed the survey. In the second time point, 113 participants completed the survey, of whom 6 respondents were excluded because they stated that their health status changed in the last two weeks, and 4 respondents were lost to follow-up, leaving 103 respondents in the reliability analysis.

The age of respondents ranged from 24 to 77 years of age, with 50.0% self-identifying as female and 50.0% as male. The majority reported having completed higher education (106 participants, 89.8%; see Table 1). The mean score of the PHQ-9 was 4 points, the minimum score was 0, and the maximum score was 19 points.

(TABLE 1)

The internal consistency of the PHQ-9 was $\alpha=0.79$ for the first time-point and 0.83 for the second time-point. We identified a good test-retest reliability ($ICC_{2,1}=0.81$), a SEM score of 1.83 points, and a MDC score of 5.07 points. The PHQ-9 moderately and negatively correlated to the SF-12's MCS (Spearman $\rho=-0.57$, $p<.05$), but not to the PCS (Spearman $\rho=-0.12$, $p=.22$; see Table 2). None of the respondents scored the highest score in the PHQ-9, while seven respondents (6.8%) scored zero.

(TABLE 2)

Short Communication

DISCUSSION

This study investigated the psychometric properties of the Hebrew version of the PHQ-9. We found that the questionnaire is reliable and valid for use among the Hebrew speaking adult population.

The internal consistency of the Hebrew version of the PHQ-9 ($\alpha=0.79-0.83$) was slightly lower than the one reported by the developers of the original, English version of the PHQ-9 ($\alpha=0.86-0.89$)⁶ and on par with other translated language versions ($\alpha=0.78-0.86$).^{10,9,7} Internal consistency describes if the items on a questionnaire measure the same general construct, and a Cronbach's alpha of 0.8 is considered acceptable.¹⁵

We identified a good test-retest reliability (ICC= 0.81), while other versions reported a reliability of 0.59-0.87. The wide range of reliability may be due to different populations (general population, HIV, breast cancer), different languages (Chinese, Portuguese, Swahili, English), different time intervals (one-eight weeks), and different sample sizes (45-187 participants).^{9,8,7} We chose a two-week time interval and 103 participants as recommended by the COSMIN guidelines.^{11,13}

The construct validity of the PHQ-9 is varied as different authors chose different comparable questionnaires. We chose the SF-12, as it was previously validated in Hebrew.¹⁴ We found the SF-12 MCS is moderately and negatively correlated with the PHQ-9 (Spearman $\rho= -0.57$, $p< .05$). Others used various SF questionnaires and reported on varying correlations from -0.43 to -0.73^{7,18,19,6} between the PHQ-9 and the mental aspect of the SF questionnaires. The differences between the studies may be due to different clinical and non-clinical populations.

Another key point is that the respondents in our study completed the PHQ-9 online. Yet, the psychometric properties of the questionnaires are on par with previous studies which administrated the survey by mail, phone, or in-person as a paper questionnaire.^{18,19,7,20} Consequently, we conclude the PHQ-9 can be administered online.

Short Communication

Our study is not without limitations. Firstly, most participants had a higher education degree, limiting the generalizability of our results to the general population. Secondly, we did not compare the PHQ-9 to a psychiatric evaluation. Thirdly, the mean PHQ-9 score of the participants was 4 points, and the highest was 19 points. Hence, people with increased depressive symptoms were underrepresented. Further studies could assess the Hebrew version of the PHQ-9 in a more varied population and compare it to a psychiatric evaluation to determine both sensitivity and specificity.

CONCLUSION

The Hebrew version of the PHQ-9 is valid and reliable for screening self-reported depressive symptoms among the Hebrew-speaking general population. Further, it is feasible to administer it online.

Short Communication

Table 1. Characteristics of the participants

	n=118
Age (years)	40.2 ± 10.8
Gender (%)	
Male	59 (50.0)
Female	59 (50.0)
Marital status (%)	
Single	14 (11.9)
In a relationship	25 (21.2)
Married	78 (66.1)
Divorced	1 (0.8)
Education level (%)	
High school/Technical school	10 (8.6)
Higher education	106 (89.8)
Missing	2 (1.7)

Table 2. Psychometric properties of the PHQ-9 questionnaire

	n=103
Cronbach's α	0.79-0.83
ICC	0.81
SEM	1.83
MDC	5.07

ICC = Intraclass Correlation Coefficient. SEM = Standard Error of Measurement. MDC = Minimal Detectable Change.

Short Communication

REFERENCES

1. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-1222.
2. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392(10159):1789-1858.
3. Lee H-J, Choi EJ, Nahm FS, Yoon IY, Lee PB. Prevalence of unrecognized depression in patients with chronic pain without a history of psychiatric diseases. *Korean J Pain*. 2018;31(2):116-124.
4. Tilahun AB, Bekele G, Mekonnen N, Tamiru E. Prevalence of unrecognized depression and associated factors among patients attending medical outpatient department in Adare Hospital, Hawassa, Ethiopia. *Neuropsychiatr Dis Treat*. 2016;12:2723-2729.
5. Fekadu A, Demissie M, Berhane R, et al. *Under Detection of Depression in Primary Care Settings in Low and Middle-Income Countries: A Systematic Review and Meta-Analysis*. 2020.
6. Kroenke K, Spitzer RL. The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatric Annals*. 2002;32(9):509-515.
7. Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2014;36(5):539-544.
8. Torres A, Monteiro S, Pereira A, Albuquerque E, eds. *Reliability and Validity of the PHQ-9 in Portuguese Women with Breast Cancer*. 2016.
9. Monahan PO, Shacham E, Reece M, et al. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *J Gen Intern Med*. 2009;24(2):189-197.
10. Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. *BMC Psychiatry*. 2008;8:46.
11. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). 2018.
12. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry*. 2000;176:249-252.

Short Communication

13. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539-549.
14. Amir M, Lewin-Epstein N, Becker G, Buskila D. Psychometric Properties of the SF-12 (Hebrew Version) in a Primary Care Population in Israel. *Medical Care.* 2002;40(10):918-928.
15. Portney LG. *Foundations of Clinical Research: Applications to Evidence-Based Practice.* Fourth. FA Davis; 2020.
16. Dancey C, Reidy J. *Statistics Without Maths for Psychology.* 7th ed. Pearson education; 2017.
17. Terwee CB, Bot SDM, Boer MR de, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.
18. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry.* 2006;28(1):71-77.
19. 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(5):551-555.
20. Pinto-Meza A, Serrano-Blanco A, Peñarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *J Gen Intern Med.* 2005;20(8):738-742.