HHS Public Access

Author manuscript

Compr Psychiatry. Author manuscript; available in PMC 2017 October 01.

Published in final edited form as:

Compr Psychiatry. 2016 October ; 70: 216–221. doi:10.1016/j.comppsych.2016.07.011.

Diagnostic Validity of the Patient Health Questionnaire-2 (PHQ-2) among Ethiopian Adults

Bizu Gelaye[a], **Ijala Wilson**[a], **Hanna Y. Berhane**[b], **Negussie Deyessa**[c], **Yonas Bahretibeb**[c], **Dawit Wondimagegn**[c], **Teshome Shibre Kelkile**[d], **Yemane Berhane**[b], **Jesse R. Fann**[e], and **Michelle A. Williams**[a]

[a]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[b]Addis Continental Institute of Public Health, Addis Ababa, ETHIOPIA

[c]Faculty of Medicine, Addis Ababa University, Addis Ababa, ETHIOPIA

[d]Horizon Health Network, Zone 3, New Brunswick, Canada

[e]Departments of Psychiatry and Behavioral Sciences, Rehabilitation Medicine, and Epidemiology, University of Washington, Seattle, WA, USA

Abstract

Background—The Patient Health Questionnaire-2 (PHQ-2) is an ultra-brief questionnaire widely used by researchers and clinicians to detect major depressive disorder (MDD). Despite its individual and societal impact, MDD is often undetected and untreated particularly among sub-Saharan Africans. We conducted this study to evaluate the reliability and validity of using the PHQ-2 as a screen for MDD among Ethiopian adults.

Methods—A total of 926 adults attending outpatient departments in a major referral hospital in Addis Ababa, Ethiopia participated in this study. Construct validity was assessed by examining associations of PHQ-2 scores with World Health Organization Quality of Life (WHO-QOL) domains. We assessed criterion validity and performance characteristics against an independent, blinded, and psychiatrist administered semi-structured Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview using measures of sensitivity, specificity and receiver operating characteristics (ROC) curves.

Results—The PHQ-2 items showed good reliability (intraclass correlation coefficient=0.92). Quality of life, as reflected by subscale scores for four WHO-QOL domains, was significantly lower among patients with increasing PHQ-2 scores demonstrating good construct validity. ROC analysis and Youden Index showed that a PHQ-2 threshold score of 3 offered optimal

Corresponding author: Dr. Bizu Gelaye, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, K505F, Boston, MA 02115 USA, Telephone: 617-432-6477, Facsimile: 617-566-7805, bgelaye@hsph.harvard.edu.

Conflict of interest: The authors have no competing interests to declare.

discriminatory power with respect to the diagnosis of MDD via the clinical interview (sensitivity=74% and specificity=60%).

Conclusion—The Amharic language version of the PHQ-2 had good sensitivity and fair specificity for detecting MDD compared against a psychiatrist administered SCAN diagnosis. This study provides evidence for the PHQ-2 as a reliable and valid ultra-brief screening tool for initial identification of MDD.

Keywords

PHQ-2; Validation; Africa; Ethiopia; Depression

Introduction

Major depressive disorder (MDD) is a major contributor to the global burden of disease and the leading cause of disability affecting some 350 million individuals worldwide [1]. Untreated MDD creates substantial health, economic and social burden for affected individuals and their families, where MDD is associated with mortality [2], cardiovascular diseases [3], HIV/AIDS [4] healthcare utilization, functional decline, and decreased quality of life [5–7]. In many low- and middle-income countries (LAMICs) MDD is highly prevalent but often under-detected and under-treated [8], in part due to the lack of skilled mental health workers, diagnostic screening instruments, as well as the stigma associated with mental illness [9–11]. There is a well-established body of evidence indicating the use of brief screening instruments for MDD, particularly in communities where limited health professionals are available. One of the most widely used MDD screening instruments is the Patient Health Questionnaire-9 (PHQ-9), a 9 item brief screener with excellent reliability and validity in cross-cultural settings [12–14].

An even shorter, two-item measure (the PHQ–2), consisting of the two cardinal symptoms of depression (depressed mood and anhedonia) has also been shown to be an excellent screening tool for MDD [15, 16]. However, few have evaluated the applicability of the PHQ-2 in LAMICs [17–19], and only three investigative teams have assessed the instrument among sub-Saharan African populations [19–21]. Therefore, the purpose of this study is to assess the validity of the PHQ-2 against a psychiatrist-administered Schedules for Clinical Assessment in Neuropsychiatry (SCAN) [22] reference/gold standard among urban dwelling Ethiopian adults. Ethiopia, with a growing population of over 84 million, is a country that is undergoing social and economic changes with increased burden of chronic diseases, particularly among urban dwelling individuals [14].

Methods and Materials

Adults (≥ 18 years of age) attending outpatient departments in St. Paul General Specialized Hospital in Addis Ababa, Ethiopia were invited to participate in the study by research nurses. Using an approach and recruitment script, a research nurse member approached and invited participants when they checked in for their appointment at the hospital. The hospital is the second largest public hospital in Ethiopia and serves as a referral and teaching hospital under the Ethiopian Federal Ministry of Health. Data collection was conducted between July

and December 2011. Study personnel included research nurses with public health training. Prior to the start of the study, research nurses were trained for four days on the contents of the questionnaire, ethical conduct of human subjects research, and data collection techniques. All study participants provided informed consent. All research protocols were approved by the Institutional Review Boards of Addis Continental Institute of Public Health, Addis Ababa, Ethiopia and the Human Subjects Division at the University of Washington, USA.

Study Procedures

We used a two-stage study design where participants were first interviewed by research nurses using nine depression screening questions from PHQ-9. Those who screened positive for depression on initial interview (i.e., positive on PHQ-9), as well as a randomly selected subgroup of participants who screened negative for depression (i.e., screen negative on PHQ-9) were invited for a diagnostic interview with a psychiatrist who was blinded to the PHQ-9 screening outcome. A total of 384 participants were invited to participate in the SCAN diagnostic interview (178 who screened positive on the PHQ-9 and 276 who screened negative on the PHQ-9) and 363 of them agreed to do so (94% of selected positive screens and 95% of selected negative screens). Those who refused to participate in the diagnostic interview cited reasons such as lack of time to do a follow-up interview with the psychiatrist. During the interview, we collected other general information including demographic characteristics, behavioral risk factors, and self-reported health status. Additionally, we collected information concerning self-reported quality of life using the WHO Quality of Life (WHO-QOL) questionnaire.

The Patient Health Questionnaire-2 (PHQ-2)

The PHQ-2 is comprised of the first two items of the PHQ-9, which is the full screening instrument for depressive disorders. As part of the original PHQ-9 validation [14], the PHQ-2, originally written in English, was translated into Amharic (the official language of Ethiopia) by the lead author. To ensure proper expression and conceptualization of terminologies in local contexts, we used a standard approach of iterative back translation by panels of bilingual experts [23, 24]. The translated version was back-translated and modified until the back-translated version was comparable with the original English version. Each question requires participants to rate the frequency of a depressive symptom experienced in the two weeks prior to evaluation. These items include: “Having little interest or pleasure in doing things” and “feeling down, depressed, or hopeless.” Scores for each item range from 0 “not at all” to 3 “nearly every day” with a total score ranging from 0 to 6. Scores of ≥3 have been recommended to identify cases positive for MDD [16].

Diagnostic Interview

The SCAN is a semi-structured clinical interview developed by the WHO for use by trained clinicians to diagnose psychiatric disorders among adults [25, 26]. The SCAN diagnostic interview, comprised of 28 modules, gives flexibility for diagnosing a number of mental disorders based on DSM-IV diagnostic criteria [25]. In this study, we used the instrument’s

depression module. The depression module has been reported to have good psychometric properties in diverse populations and in multiple languages [27–29]. Of note, the Amharic version of the SCAN including the depression module was shown to be a feasible and reliable tool in Ethiopia (percent agreement 93.0 with kappa=0.80) among Ethiopians [27]. On the basis of earlier reports, we elected to use the psychiatrist administered interview (using SCAN) as the reference (or gold) standard against which we would test the reliability and validity of the PHQ-2 as a diagnostic and screening questionnaire for MDD among Ethiopians.

Statistical Analyses

First, we computed interclass correlation coefficients to assess the test-retest reliability of PHQ-2 instrument among 5% of the study participants who completed two questionnaires on the same day. Next, we used the WHO-QOL questionnaire to evaluate the construct validity of the PHQ-2 instrument. The WHO-QOL is a cross-cultural assessment tool that captures an individual's perception of their position in life in the context of culture and value systems in which they live and in relation to their goals, expectations, standards and concerns [30]. The instrument has been widely used globally including in sub-Saharan Africa [30]. This study used an abbreviated version of WHO-QOL (also known as WHOQOL-BREF) which has 26 items that cover four domains: physical health, psychological health, social relationships, and environment. The overall percentile score for each domain ranges from 0% (very poor) to 100% (very good). Since prior research has shown statistically significant associations between depression and quality of life [31, 32], we hypothesized that lower WHO-QOL scores would be associated with higher PHQ-2 scores.

Convergent validity was measured using Pearson's correlation between the PHQ-2 and PHQ-9. Finally, we assessed the criterion validity by determining the concordance between the PHQ-2 score and a psychiatrist diagnosis of MDD using the SCAN. Parameters computed were: sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive values, and negative predictive values for the presence or absence of MDD. Additionally, to identify the best PHQ-2 cut-off score to use in depression screening, we completed receiver operating characteristic (ROC) curve analyses to identify optimal balance of sensitivity and specificity, area under the ROC curve (AUC) and its nonparametric 95% CI [33]. Additionally, we calculated the Youden Index, a metric used for identifying an optimum cut-off score for screening test [34]. The Youden index is a function of sensitivity and specificity calculated as the (sensitivity + specificity − 1). The range of the index is from 0 to 1 with the higher values designating an optimum cut-of[34]f.

Results

A summary of selected socio-demographic and lifestyle characteristic of the study sample is presented in Table 1. A total of 926 participants between the ages of 18 and 69 years (mean age=35 years, standard deviation (SD)±11.7 years) participated in the study. The majority of participants were women (61%), married (52.3%) and with less than high school education (78%). Approximately 4% of participants reported that they were current smokers, 5.3 %

chewed khat and 9.6% reported consuming at least one alcoholic beverage per week. Nearly 44% of participants reported having a fair or poor physical health status, while 33% reported poor mental health status. Selected socio-demographic and lifestyle characteristics of those participants with SCAN interview (N=363) is also presented in Table 1.

As shown in Table 2, across all domains, mean WHO-QOL scores decreased with increasing PHQ-2 total score. The highest correlation between PHQ-2 and WHO-QOL was observed for psychological domain ($r=-0.51$, $p<0.01$) with the other domains having similar correlations of about $r=0.35$ (p-value<0.01). In addition, higher levels of depressive symptom severity using the PHQ-2 scores were highly correlated with PHQ-9 scores ($r=0.80$, $p<0.001$) demonstrating convergent validity. However, due to the two identical items, this correlation coefficient might be especially high. The test-retest reliability intraclass correlation coefficient of the PHQ-9 total score was 0.92.

We next evaluated the psychometric properties of the PHQ-2 using the SCAN diagnosis as a gold standard (Table 3). The optimal cut point for maximizing the sensitivity of the PHQ-2 without loss of specificity was a score of ≥3. At this cut point, the PHQ-2 had a sensitivity of 73.9% (95% CI: 61.3–72.9%) and a specificity of 59.6% (95% CI: 61.3–72.9%). The area under the ROC curve (AUC) for detecting MDD at this cut off was 0.72 (95% CI: 0.64–0.79) (Supplemental Figure 1).

Discussion

Our results provided evidence that the Amharic version of the PHQ-2 is a tool with good construct validity and high reliability (intraclass correlation coefficient=0.92) among Ethiopian adults. Lower quality of life scores across physical, psychological, social relationship and environmental domains were noted with increasing PHQ-2 scores demonstrating good construct validity of PHQ-2. ROC analysis and Youden Index showed that compared to psychiatrist-administered SCAN reference standard, the PHQ-2 had good sensitivity (74%) and fair specificity (60%) with respect to diagnosis of MDD at a cut-off score of ≥3. However, other cut-off scores might be used if the main emphasis of a study is for either higher sensitivity or higher specificity. In a clinical settings, a good screening instrument should have a high sensitivity with an acceptable specificity in order to avoid missing MDD cases.

The PHQ-2 was most successful at identifying a person without MDD (with a good negative predictive value of 94%). In other words, a person who was identified as non-depressed using PHQ-2 had 94% probability of not having MDD diagnosis using SCAN. However, a person who was identified as depressed using PHQ-2 had 21% (PPV) chance of having MDD diagnosis using SCAN. The low PPV might due to the low prevalence of MDD in our study population. The positive likelihood ratio commonly considered to “rule in disease” of the PHQ-2 was 1.8. The negative LR commonly considered to “rule out of disease” was 0.6. This means that clinically in a similar outpatient setting, patients with MDD are 1.8-times more likely to have a PHQ-2 score ≥ 3 compared to patients without MDD. Similarly patients without MDD are 1.7 (1/0.6) times more likely to have negative test results using PHQ-2 compared to those with MDD.

Kroenke et al, in their original validation study of PHQ-2, found that at a cut point of 3, the PHQ-2 has a sensitivity of 83% and a specificity of 90% [16]. The performance characteristics reported in our study are lower than the original validation study [16] but higher than prior studies conducted among other sub-Saharan African populations [20, 21]. Bhana et al, in their study among chronic care patients attending primary health facilities in South Africa, found a sensitivity of 39.0% and as specificity of 91.7% [21] while a better sensitivity (80.5%) and satisfactory specificity (57.5) was noted at a cutoff of 1. In their study among rural residents of Butajira, Ethiopia, Hanlon et al found a sensitivity of 33% and specificity of 93.1% at a cutoff of 3 and sensitivity of 83.3% and specificity of 60.8% at a cutoff of 1 [20]. Finally, Monahan et al among HIV/AIDS patients in Western Kenya found the PHQ-2 to have good construct validity and reliability for assessing MDD although the criterion validity was not established in this study using a gold standard [19]. Despite differences in population characteristics, sample size, and study settings, on balance, the findings of our study and those of others document the potential benefits of using the PHQ-2 as a first screening tool among sub-Saharan Africans [16, 19–21].

Consistent with the WHO Mental Health Gap Action Programme (mhGAP), the Ethiopian Federal Ministry of Health recently developed a National Mental Health Strategy that outlines decentralized and integrating mental health services into primary health care level [35]. As part of this national strategy, brief screening instruments for common mental health disorders are particularly needed in primary health care settings. The integration of mental health in to primary health care is one of the widely suggested strategies to adequately address the burden of mental disorders in LAMICs [36]. Our study findings provide evidence for utility of the PHQ-2 questionnaire as a first screening tool for the early detection of MDD particularity in such an integrated mental health service system where administration of the PHQ-9 is feasible. The PHQ-9 would still be the preferred instrument to diagnoses MDD, monitor depression severity and evaluate patients' response to treatment [14] as we have demonstrated its reliability and validity for MDD diagnosis among Ethiopian adults [14]. As a clinical approach in this study population sequential screening where the PHQ-2 can be used as an initial screener followed by the PHQ-9 if the results of the PHQ-2 are positive might be a good strategy for detecting MDD.

There are some caveats that must be considered when interpreting the results of our study. First, our study was conducted in an urban hospital; therefore, the result may not be generalizable to populations in rural and remote areas. Second, since the PHQ-2 questionnaire and the SCAN diagnostic interview were conducted during the same day, it is possible that the short time interval between the two interviews administration created carryover or recall effects and increased the reliability [37]. The consistency of our study findings with those of other studies [16, 19–21] and the psychometric properties of other reliability measures, in part, mitigate this concern. Finally depression screening alone is insufficient for addressing growing mental health care needs in LAMICs settings although having a valid screening instrument is an important first step towards addressing them.

In summary our study provides evidence to support the use of an Amharic version of PHQ-2 for depression screening. The strong construct and good criterion validity of the PHQ-2 make it an attractive ultra-brief screening questionnaire to be used widely at the primary

health care level. This finding is particularly important given that studies conducted in sub-Saharan Africa among primary health clinic patients show that 20–30% of such patients present with MDD and other psychiatric disorders as the primary or secondary reason for seeking medical care [38]. Access to and systematic use of ultra-brief screeners in primary care settings simplifies depression screening and enhances routine use in clinical settings where mental health care is integrated. The use of ultra-brief screeners such as the PHQ-2 is only the first step in identifying patients with MDD for further investigation and initiation of appropriate treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by an award from the National Institutes of Health, National Institute of Minority Health and Health Disparities (T37-MD001449). The authors wish to thank the staff of Addis Continental Institute of Public Health for their expert technical assistance. The authors would also like to thank Saint Paul Hospital for granting access to conduct the study

References

1. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson FJ, Norman RE, Flaxman AD, Johns N, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet. 2013; 382(9904):1575–1586. [PubMed: 23993280]
2. Park M, Katon WJ, Wolf FM. Depression and risk of mortality in individuals with diabetes: a meta-analysis and systematic review. Gen Hosp Psychiatry. 2013; 35(3):217–225. [PubMed: 23415577]
3. Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J. 2006; 27(23):2763–2774. [PubMed: 17082208]
4. Owe-Larsson B, Sall L, Salamon E, Allgulander C. HIV infection and psychiatric illness. Afr J Psychiatry (Johannesbg). 2009; 12(2):115–128. [PubMed: 19582313]
5. Simon G, Ormel J, VonKorff M, Barlow W. Health care costs associated with depressive and anxiety disorders in primary care. Am J Psychiatry. 1995; 152(3):352–357. [PubMed: 7864259]
6. Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA. 1990; 264(19):2524–2528. [PubMed: 2146410]
7. Lowe B, Spitzer RL, Grafe K, Kroenke K, Quenter A, Zipfel S, Buchholz C, Witte S, Herzog W. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. J Affect Disord. 2004; 78(2):131–140. [PubMed: 14706723]
8. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, Angermeyer MC, Bernert S, de Girolamo G, Morosini P, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA. 2004; 291(21):2581–2590. [PubMed: 15173149]
9. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, Simon G, Weiss HA. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. Psychol Med. 2008; 38(2):221–228. [PubMed: 18047768]
10. Wig NN. WHO and mental health–a view from developing countries. Bull World Health Organ. 2000; 78(4):502–503. [PubMed: 10885170]
11. Patel V, Sartorius N. From science to action: the Lancet Series on Global Mental Health. Curr Opin Psychiatry. 2008; 21(2):109–113. [PubMed: 18332652]

12. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001; 16(9):606–613. [PubMed: 11556941]
13. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. Am J Obstet Gynecol. 2000; 183(3):759–769. [PubMed: 10992206]
14. Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, Wondimagegn D, Lemenhe A, Fann JR, Vander Stoep A, et al. Validity of the Patient Health Questionnaire-9 for depression screening and diagnosis in East Africa. Psychiatry Res. 2013; 210(2):653–661. [PubMed: 23972787]
15. Lowe B, Kroenke K, Grafe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res. 2005; 58(2):163–171. [PubMed: 15820844]
16. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care. 2003; 41(11):1284–1292. [PubMed: 14583691]
17. Li C, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. J Am Geriatr Soc. 2007; 55(4):596–602. [PubMed: 17397440]
18. Chagas MH, Crippa JA, Loureiro SR, Hallak JE, Meneses-Gaya C, Machado-de-Sousa JP, Rodrigues GR, Filho AS, Sanches RF, Tumas V. Validity of the PHQ-2 for the screening of major depression in Parkinson's disease: two questions and one important answer. Aging Ment Health. 2011; 15(7):838–843. [PubMed: 21562987]
19. Monahan PO, Shacham E, Reece M, Kroenke K, Ong'or WO, Omollo O, Yebei VN, Ojwang C. 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. [PubMed: 19031037]
20. Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, Lund C, Prince M, Fekadu A. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. J Affect Disord. 2015; 186:32–39. [PubMed: 26226431]
21. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. BMC Psychiatry. 2015; 15:118. [PubMed: 26001915]
22. Rashid E, Kebede D, Alem A. Evaluation of an Amharic version of the Composite International Diagnostic Interview (CIDI) in Ethiopia. Ethiop J Health Dev. 1996; 10:69–77.
23. Kinder LS, Carnethon MR, Palaniappan LP, King AC, Fortmann SP. Depression and the metabolic syndrome in young adults: findings from the Third National Health and Nutrition Examination Survey. Psychosom Med. 2004; 66(3):316–322. [PubMed: 15184689]
24. Kessler, RC.; Ustün, TB.; World Health, O. The WHO World Mental Health Surveys: global perspectives on the epidemiology of mental disorders. Cambridge; New York; Geneva: Cambridge University Press; Published in collaboration with the World Health Organization; 2008.
25. Aboraya A, Tien A, Stevenson J, Crosby K. Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community. W V Med J. 1998; 94(6):326–328. [PubMed: 9868378]
26. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry. 1990; 47(6): 589–593. [PubMed: 2190539]
27. Alem A, Kebede D, Shibre T, Negash A, Deyassa N. Comparison of computer assisted scan diagnoses and clinical diagnoses of major mental disorders in Butajira, rural Ethiopia. Ethiop Med J. 2004; 42(2):137–143. [PubMed: 16895030]
28. Cheng AT, Tien AY, Chang CJ, Brugha TS, Cooper JE, Lee CS, Compton W, Liu CY, Yu WY, Chen HM. Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. Br J Psychiatry. 2001; 178:567–572. [PubMed: 11388976]
29. Krisanaprakornkit T, Paholpak S, Piyavhatkul N. The validity and reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Mood Disorders Section. J Med Assoc Thai. 2006; 89(2):205–211. [PubMed: 16579007]

30. Saxena S, Carlson D, Billington R, Life WGWHOQO. The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation. 2001; 10(8):711–721.
31. Hyphantis T, Kotsis K, Voulgari PV, Tsifetaki N, Creed F, Drosos AA. Diagnostic accuracy, internal consistency and convergent validity of the Greek version of PHQ-9 in diagnosing depression in rheumatological disorders. Arthritis Care Res (Hoboken). 2011
32. Omoro SA, Fann JR, Weymuller EA, Macharia IM, Yueh B. Swahili translation and validation of the Patient Health Questionnaire-9 depression scale in the Kenyan head and neck cancer patient population. Int J Psychiatry Med. 2006; 36(3):367–381. [PubMed: 17236703]
33. Zhou, XH.; McClish, DK.; Obuchowski, NA. Statistical Methods in Diagnostic Medicine. Wiley-Blackwell; Oxford: 2011.
34. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1):32–35. [PubMed: 15405679]
35. EMOH. Federal Democratic Republic of Ethiopia Ministry of Health National Mental Health Strategy Addis Ababa, Ethiopia. 2012
36. Patel V. Mental health in low- and middle-income countries. Br Med Bull. 2007:81–82. 81–96.
37. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol. 2003; 56(8):730–735. [PubMed: 12954464]
38. Ngoma MC, Prince M, Mann A. Common mental disorders among those attending primary health clinics and traditional healers in urban Tanzania. Br J Psychiatry. 2003; 183:349–355. [PubMed: 14519614]

Table 1

Characteristics of the study population

Characteristic	N=926		Depressed N=46	Non-depressed N=317	P-value
	n	%	%		
Age (years), Mean± SD	35.1±1.7		33.7 ± 9.6	35.1 ± 11.9	0.448
Sex					
Women	568	61.3	80.4	60.6	0.001
Men	358	38.7	19.6	39.4	
Marital status					
Married	486	52.5	36.9	53.3	<0.001
Never married	293	31.6	28.3	30.3	
Other	147	15.9	34.8	16.4	
Education					
≤ Primary (1–6)	400	43.2	52.2	45.7	0.425
Secondary (7–12)	322	34.8	30.4	34.7	
College graduate	204	22.0	17.4	19.6	
Smoking status					
Never	802	86.6	78.3	86.4	0.329
Former	88	9.5	17.4	10.4	
Current	36	3.9	4.3	3.2	
Alcohol consumption past year					
Non-drinker	528	57.0	73.9	55.2	0.060
Less than once a month	309	33.4	21.7	34.4	
≥ 1 day a week	89	9.6	4.4	10.4	
Khat consumption					
None	679	73.7	69.6	72.3	0.623
Former	198	21.4	2.2	3.8	
Current	49	5.3	28.2	23.9	
Self-reported physical health					
Excellent/very good/good	522	56.4	36.9	47.9	0.002
Poor/fair	404	43.6	63.1	52.1	

Characteristic	N=926		Depressed N=46	Non-depressed N=317	P-value
	n	%	%		
Self-reported mental health					
Excellent/very good/good	616	66.5	34.8	56.2	0.007
Poor/fair	310	33.5	65.2	43.8	

*The SCAN interview as conducted among 363 participants

Table 2

Construct validity: association between PHQ-2, PHQ-9 and WHO quality of life scores (N=926)

Level of depressive symptom severity	PHQ-9 depression		Quality of life by domain							
	Mean		Physical Mean score		Psychological Mean score		Social relationships Mean score		Environmental Mean score	
0 (n= 272)	1.9		57.9		68.1		72.8		49.5	
1 (n=181)	4.2		52.9		58.8		68.3		43.8	
2 (n=189)	7.4		48.0		50.7		59.5		38.9	
3 (n=106)	9.8		49.0		48.9		58.1		38.4	
4 (n=94)	11.7		45.3		45.2		51.5		36.5	
5 (n=27)	13.2		45.2		41.5		52.9		34.7	
6 (n=57)	15.8		41.9		38.4		47.2		31.3	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
Correlation with PHQ-2	0.80	<0.001	−0.35	<0.001	−0.51	<0.001	−0.35	<0.001	−0.34	<0.001

Table 3

Sensitivity and specificity for MDD diagnosis across various cut-off points of the PHQ-2 (N=363)

PHQ-2 cut off score	Sensitivity (95% CI)	Specificity (95% CI)	Youden Index	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR− (95% CI)
≥1	97.8 (88.5–99.9)	15.8 (11.9–20.3)	13.0	14.4 (10.7–18.8)	98.0 (89.6–99.9)	1.2 (1.1–1.2)	0.2 (0.1–0.6)
≥2	91.3 (79.2–97.6)	35.3 (30.1–40.9)	26.6	17.0 (12.5–22.3)	96.6 (91.4–99.1)	1.4 (1.3–1.6)	0.4 (0.3–0.7)
≥3	**73.9 (58.9–85.7)**	**59.6 (54.0–65.1)**	**33.6**	**21.0 (15.0–28.1)**	**94.0 (89.8–96.9)**	**1.8 (1.5–2.3)**	**0.6 (0.4–0.8)**
≥4	54.3 (39.0–69.1)	75.7 (70.6–80.3)	30.0	24.5 (16.5–34.0)	92.0 (88.0–95.0)	2.2 (1.6–3.1)	0.8 (0.6–1.0)
≥5	30.4 (17.7–45.8)	88.6 (84.6–91.9)	19.0	28.0 (16.2–42.5)	89.8 (85.9–92.9)	2.7 (1.6–4.6)	0.9 (0.7–1.0)
≥6	21.7 (10.9–36.4)	91.2 (87.5–94.1)	12.9	26.3 (13.4–43.1)	88.9 (85.0–92.1)	2.5 (1.3–4.7)	1.0 (1.0–1.0)

PPV: positive predictive value, NPV−:negative predictive value; LR+: positive likelihood ratio, LR−: negative likelihood ratio