Reliability and Validity of a Modified PHQ-9 Item Inventory (PHQ-12) as a Screening Instrument for Assessing Depression in Asian Indians (CURES - 65)

S Poongothai, R Pradeepa, A Ganesan, V Mohan

Abstract

Objectives: To evaluate the validity and reliability of the modified Patient Health Questionnaire (PHQ) 12 item instrument as a screening tool for assessing depression compared to the PHQ-9 in a representative south Indian urban population.

Methods: The Chennai Urban Rural Epidemiology Study [CURES] is a large cross-sectional study conducted in Chennai, South India. In Phase 1 of CURES (urban component), 26,001 individuals aged ≥ 20 years individuals were selected by a systematic sampling technique of whom one hundred subjects were randomly selected, using computer-generated numbers, for this validation study. Two self-reported questionnaires (modified PHQ-12 item and PHQ-9 item) were administered to the subjects to compare their effectiveness in detecting depression. Reliability and validity were assessed and Receiver Operating Characteristic (ROC) curves were plotted. Pearson’s correlation was used to compare the two questionnaires.

Results: The mean age of the study was 38.6±11.6 years and 48% were males. Pearson’s correlation coefficient between the modified PHQ-12 and the PHQ-9 item was 0.913 [p<0.0001]. Factor Analysis revealed that the modified PHQ-12 item scale can be used as a unidimensional scale and had excellent internal consistency (Cronbach’s alpha: 0.88). A cut point of >4 calculated using the ROC curves for the modified PHQ-12 item had the highest sensitivity (92.0%) and specificity (90.7%) using PHQ-9 as the gold standard. The positive predictive value was 76.7%, and the negative predictive value, 97.1% and the area under the ROC curve, 0.979 (95% Confidence Interval: 0.929 - 0.997, p<0.0001).

Conclusion: The modified PHQ-12 item is a valid and reliable instrument for large scale population based screening of depression in Asian Indians and a cut point score of greater than 4 gave the highest sensitivity and specificity. ©

INTRODUCTION

The prevalence of non communicable diseases (NCD’s) are rapidly rising in India. The term NCD’s not only includes diabetes, hypertension, obesity and cardiovascular disease but also cancers, chronic respiratory disease, injuries and mental illnesses. There are very few population based studies on the prevalence of depression in India and there is an obvious urgent need for such data. Among western countries, the prevalence of depressive disorders is fairly high in the general population and even higher in primary care and general hospital settings. Depression is associated with severe impairment in physical and social functioning, leading to higher health care utilization and costs.1,2 Depression in primary care is common, disabling and costly but treatable if identified early by screening.3,4 Patients suffering from depressive disorders often do not seek help for psychological problems, but instead present with somatic symptoms to their physicians and hence their depression often goes unrecognized.3,5 According to the WHO Psychological Problems in General Health Care study, only 42% of primary care patients with major depression were recognized by their physicians as having depression.7 Therefore it is clear that a suitable screening tool has to be used to aid in the early detection of depression and thus facilitate clinical decision making.

There are different tools to measure the levels of depression both in the primary care and in the general population. Assessment of depression is commonly performed using generic depression rating scales with semi structured Hamilton rating scale for depression (HAM-D)6 or with self-reported measures such as the Beck Depression Inventory (BDI),9 Centre for Epidemiologic Studies Depression Scale (CES-D),10 Beck Hopelessness...
A number of studies on the validity and reliability of PHQ-9, as a diagnostic measure as well as its utility in assessing depression severity and in monitoring treatment responses have been published in western countries. However, our knowledge, no study has examined the validity of the PHQ-9 in a south Asian and particularly Asian Indian population. We further modified the PHQ-9 and developed a PHQ-12 questionnaire suitable for south Asian population and validated this against the conventional PHQ-9, as a possible screening tool for detecting depression in large scale population-based studies. To our knowledge, this is the first study to validate the use of PHQ-9 and PHQ-12 instruments in India and is significant given the need for population-based studies on the prevalence of depression in India.

**METHODS**

The subjects for this study were recruited from the urban component of the Chennai Urban Rural epidemiological Study (CURES), conducted on a representative population of Chennai, the methodology of which has been published elsewhere. Briefly, the city of Chennai is divided into 155 corporation wards representing a socio-economically diverse group. In Phase 1 of CURES, 26,001 individuals aged ≥ 20 years individuals from 46 corporation wards were screened using a systematic sampling technique. From this Phase, we randomly selected, using computer-generated numbers, 100 unrelated subjects and invited them to participate in the present study prior to the commencement of the main study.

Both the self-reported questionnaires (the modified PHQ-12 item and PHQ-9 item) were administered to all subjects, after making them sit in a comfortable position. The process was explained to the subjects in detail and a written informed consent was obtained from the study participants. Each questionnaire took 10 minutes to administer and was done a week apart. The study was approved by the Institutional Ethics Committee of the Madras Diabetes Research Foundation.

**INSTRUMENTS**

**Patient Health Questionnaire (PHQ-9 item)**

The PHQ-9 is a 9-item self-reported questionnaire designed to evaluate the presence of depressive symptoms during the prior 2 weeks. The nine items of the PHQ-9 are based directly on the nine diagnostic criteria for major depressive disorder in the DSM-IV (Diagnostic and Statistical Manual Fourth Edition). This can help track a patient's overall improvement or not, with treatment. As a severity measure, scores range from 0 (absence of depressive symptoms) to 27 (most severe depressive symptoms). Each of the 9 items, by asking for each of the DSM-IV symptoms, can be scored from 0 (not at all) to 3 (nearly every day). As a diagnostic measure, major depression is diagnosed if 5 or more of the 9 depressive symptoms have been present for at least "more than half, the days" (ie a score of 2) during the past 2 weeks, and if one of the symptoms is depressed mood. In addition, before making a final diagnosis, causes of acute depression due to a recent physical problem, bereavement and/or a history of manic disorder, have to be ruled out.

**Modified Patient Health Questionnaire (PHQ-12 item)**

The modified PHQ-12 item is very similar to the brief Patient Health Questionnaire Mood Scale (PHQ-9) which was derived from the Primary Care Evaluation of Mental Disorders (PRIME-MD) except for the modifications outlined below:

Three of the questions in the PHQ-9 item were split into 2 questions in order to make it more user friendly. The response categories were also modified further by making the replies made dichotomous (yes/no), such that the patient would be asked whether they felt each of the twelve depressive symptoms and their frequency during the last two weeks. The reason for making the response as yes or no in the modified PHQ-12 item is because it can be used with ease to screen for depression in large epidemiological studies. As we modified the original PHQ-9 item, validation of the new questionnaire became necessary. Table 1 compares the PHQ-9 and the modified PHQ-12 questionnaires.

Statistical Analysis

Statistical analysis was done using SAS statistical package (version 9.0; SAS Institute, Inc., Cary, NC). Numbers are expressed as mean ± SD. Pearson's correlation coefficient was calculated to determine the correlation between the modified PHQ-12 item and the PHQ-9 item as a measure of convergent validity of the PHQ-12. Receiver Operating Characteristics (ROC) curves, which are plots of the sensitivity versus 1-specificity were constructed to determine the optimal threshold point for the modified PHQ-12 item by using the PHQ-9 item as the gold standard to detect depression. The sensitivity of a specific threshold point was defined as its ability to correctly identify individuals with depression, and its specificity was defined as its ability to correctly identify individuals without depression. The area under the ROC curve (AUC) was obtained. The positive and negative predictive values for identifying subjects with depression were also calculated. The internal consistency of the modified PHQ-12 item was evaluated using Cronbach's
RESULTS

A total of 100 subjects were randomly selected from CURES and all the 100 subjects completed both the modified PHQ-12 item and PHQ-9 item (100% response rate). The mean age of the study group was 38.6 ± 11.6 years and 48% were males (Table 2).

The Pearson’s correlation coefficient between the PHQ-12 and the PHQ-9 item was 0.913 [p<.0001]. The excellent correlation between the two tools indicates that the modified PHQ-12 item measures a similar construct to that assessed by the PHQ-9 item.

Figure 1 depicts a scatter-plot illustrating the linear association between PHQ-9 Item scores and PHQ-12 Item scores with an $r^2 = 0.84$ showing that the PHQ-12 item is very sensitive and it captures even very mild levels of depression. It is thus an ideal tool for detecting any form or degree of depression.

Receiver Operator Characteristics were computed to evaluate the effectiveness of the modified PHQ-12 item scale in predicting depression by using the PHQ-9 item as the gold standard. Fig. 2 presents the ROC and the specificity and sensitivity of the different cut-off scores of the modified PHQ-12 item scale to detect depression. At a cut-off score of greater than 4 gave the higher sensitivity (92%) and specificity (90.7%). This threshold had a positive predictive value (PPV) of 76.7% and a negative predictive value (NPV) of 97.1%. The area under the ROC curve was 0.979 (95% confidence interval (CI): 0.929 to 0.997, p<0.0001).

Internal consistency

The internal consistency of the modified PHQ-12 item was evaluated using Cronbach’s $\alpha$ test. Factor analyses were also conducted on the modified PHQ-12 items in order to examine their factor structures.

Factor analysis

Earlier studies from the west have demonstrated the usefulness and validity of the PHQ-9 to measure depression and suggested a cut off score for the diagnosis of depression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>38.6 ± 11.6</td>
</tr>
<tr>
<td>Men (%)</td>
<td>48 %</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>112 ± 14</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>69 ± 5</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>52 ± 11</td>
</tr>
<tr>
<td>Height (cms)</td>
<td>155 ± 8</td>
</tr>
<tr>
<td>Body Mass Index (kg/m²)</td>
<td>21.8 ± 4.7</td>
</tr>
<tr>
<td>Socio economic status (%)</td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>32.6</td>
</tr>
<tr>
<td>Middle income</td>
<td>45.1</td>
</tr>
<tr>
<td>High income</td>
<td>22.3</td>
</tr>
<tr>
<td>Marital status (Yes) (%)</td>
<td>88</td>
</tr>
<tr>
<td>Education status (%)</td>
<td></td>
</tr>
<tr>
<td>Illiterate</td>
<td>30.2</td>
</tr>
<tr>
<td>Below SSC</td>
<td>16.2</td>
</tr>
<tr>
<td>Only SSC</td>
<td>12.3</td>
</tr>
<tr>
<td>Graduate</td>
<td>19.5</td>
</tr>
<tr>
<td>Post graduate</td>
<td>16.8</td>
</tr>
<tr>
<td>Professional</td>
<td>5.0</td>
</tr>
</tbody>
</table>

DISCUSSION

Earlier studies from the west have demonstrated the usefulness and validity of the PHQ-9 to measure depression and suggested a cut off score for the diagnosis of depression.
in the general population,\textsuperscript{13} in primary health care\textsuperscript{18} and in medically ill patients.\textsuperscript{19} The purpose of this study was to evaluate the validity and reliability of a modified version of PHQ-9 [PHQ-12 item] in a representative urban Asian Indian population. We report that the PHQ-12 item depression tool is a simple instrument that can be either administered or self reported to assess depression in epidemiological studies. The PHQ-12 item was found to have good test–retest reliability and internal consistency in comparison to PHQ-9 item tool and a score of greater than 4 appears to be the optimal cut off for screening for depression in this population. It is also sensitive to pick up even mild cases of depression and also identify the behavioural changes. This is the first study to our knowledge to validate this depression screening instrument in our country.

The PHQ-9 is a relatively short, self-administered scale designed to be a 1-step instrument for the diagnosis of depression and for the assessment of its severity. This is one of the most distinct advantages of the PHQ-9 instrument; thus, its use has been advocated to assist in the detection of depression and for follow-up to detect symptom changes and their improvement with treatment.\textsuperscript{20} The use of the 2-item\textsuperscript{14} and 15-item\textsuperscript{21} version of the PHQ has also been validated in clinic populations in the west. This study validates the PHQ-12 item in a population based study in India and thus it is even more significant. The PHQ-12 item is very easy to score as it contains only 12 questions. It has comparable psychometric properties to the longer versions and takes only around ten minutes to complete. Given its speed of administration, it is well suited for use in large epidemiological research studies where it is impractical to administer more complicated instruments to assess depression by psychiatric specialists.

The validity of the PHQ-12 in our study was supported by the excellent AUC [0.979] determined by ROC analysis, which incorporates both sensitivity and specificity to estimate the probability that a scale will correctly classify individuals.\textsuperscript{22} Values greater than 0.50 indicate better than chance classification and an AUC of 0.80 or higher suggest that the instrument is useful\textsuperscript{23} in screening for depression. Using a cut-off value of greater than 4, the sensitivity was 0.92 and specificity was 0.91 in our study. These values of sensitivity and specificity are highly acceptable as suggested by Robins et al.\textsuperscript{24} Indeed, a screening tool is considered reliable when the sensitivity range is 0.79 and the specificity > 0.63.\textsuperscript{24}

Streiner et al,\textsuperscript{25} have suggested that for a self-report instrument to be reliable and internally consistent suggested that Cronbach’s alpha test be at least 0.70. The internal consistency of the PHQ-12 in this study using cronbacha’s alpha was 0.86 which suggests that the PHQ-12 item instrument is a highly reliable and a valid tool for screening for depression in our population.

One of the advantages of the PHQ-12 is its exclusive focus on the diagnostic criteria for DSM-IV depressive disorders. Furthermore the PHQ-12 item incorporates symptoms similar to PHQ-9 item which are not found in the original DSM-IV criteria (e.g. loneliness, hopelessness,
acknowledge gratefully the help of Dr. Latha Sathish, Assistant Professor, Department of Psychology, University of Madras, Chennai for her help with this study. We thank Ms.K.Karkuzhali, Ms.Gayathiri, Ms.Vetrivelai and Ms.Sivagamasundari from MDRF for the fieldwork and most importantly the subjects who participated in the study. We are grateful to Chennai Willingdon Corporate Foundation, Chennai, for the financial support provided for the study. This is the 65th paper from the CURES study (CURES-65).

REFERENCES

22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiving Operating characteristic (ROC) curve. Radiology


---

Announcement

Books Edited by Dr. PC Manoria

A) Cardio-Metabolic Medicine : Emerging Frontiers 2008 : Rs. 300.00
B) Cardio-Diabetology Update-2007 : Rs. 300.00
C) Echocardiography update edited by Dr. PC Manoria
   Volume I : Valvular heart disease : Rs. 400.00
   Volume II : Preclinical atherosclerosis and CAD : Rs. 400.00
   Volume III : Congenital heart disease : Rs. 400.00
   Volume IV : Endocardial, myocardial, pericardial disease and cardiac tumours : Rs. 400.00
   Volume V : Normal echocardiography : Rs. 400.00
D) Cardiology update edited by Dr. PC Manoria
   Volume I : Drugs for heart and diabetes : Rs. 300.00
   Volume II : Cardiac emergencies : Rs. 300.00

For procuring above books please contact : Dr. PC Manoria
Ph. No. : 0755-2422299, 9393042229
E-mail : pmanoria@rediffmail.com
Address : E-5/103 Arera Colony, Bhopal - 462003