Original Article

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 4 or greater 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]

Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre, Gopalapuram, Chennai, India.
Received :?; Accepted : ?

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.[5,6] 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)[8] or with self-reported measures such as the

Beck Depression Inventory (BDI),[9] Centre for Epidemiologic Studies Depression Scale (CES-D),[10] Beck Hopelessness Scale and Geriatric Depression Scale(GDS)[11] and the Patient Health Questionnaire-9 (PHQ-9).[12] The PHQ-9 is a 9-item self-reported questionnaire designed to evaluate the presence of depressive symptoms during the previous 2 weeks and is effectively used to measure depression both in the clinic and the general population.[13] A shorter version of the PHQ-9 item is also used to measure depression which consists of just two items and is known as Patient Health Questionnaire (PHQ-2 item) with a yes or no response but it is recommended that even if PHQ-2 item is positive, the PHQ-9 should be additionally used.[14]

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.[11,13,15] However to 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.[16] 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 Heath 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 depression severity as whether the specific symptom(s) are improving or not, with treatment. As a severity measure, scores can 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.[14]

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) [17] 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 in Indian conditions and it thus became a 12 item questionnaire. 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 scale as a measure of convergent validity of the PHQ-12. Receiver Operating Characteristics (ROC) curves, which are plots of the sensitivity versus 1-

Table 1 : Comparison of the PHQ-9 and the modified PHQ-12 questionnaires

Patient Health Questionnaire (PHQ) -9 item	Modified Patient Health Questionnaire (PHQ) -12 item
Response: Not at all/Several days/More than half the days/Nearly every day	Response: Yes/No
1. Little interest or pleasure in doing things	1. Feeling sad, blue or depressed
2. Feeling down, depressed, or hopeless	2. Loss of interest or pleasure in most things
3. Trouble falling asleep, staying asleep,or sleeping too much	3. Feeling tired or low on energy most of the time
4. Feeling tired or having little energy	4. Loss of appetite or weight loss
5. Poor appetite or overeating	5. Overeating or weight gain
6. Feeling bad about yourself, feeling that you are a failure, or feeling that you have let yourself or your family down	6. Trouble falling asleep or staying asleep
7. Trouble concentrating on things such as reading the newspaper or watching television	7. Sleeping too much
8. Moving or speaking so slowly that other people coul have noticed. Or being so fidgety or restless that you have been moving around a lot more than usual	8. More trouble than usual concentrating on things
9. Thinking that you would be better off dead or that you want to hurt yourself in some way	9. Feeling down on yourself, no good, or worthless
	10. Being fidgety or restless that you move around a lot more than usual
	11. Moved or spoke so slowly that other people could have noticed
	12. Thought about death more than usual, either your own, someone else's, or death in general

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 ± test. Factor analyses were also conducted on the modified PHQ-12 items in order to examine their factor structures.

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).

Table 2 : General characteristics of the study population (n = 100)

Variables	
Age (years)	38.6 ± 11.6
Men (%)	48 %
Systolic blood pressure (mm Hg)	112 ± 14
Diastolic blood pressure (mm Hg)	69 ± 5
Weight (kg)	52 ± 11
Height (cms)	155 ± 8
Body Mass Index (kg/m^2)	21.8 ± 4.7
Socio economic status (%)	
Low income	32.6
Middle income	45.1
High income	22.3
Marital status (Yes) (%)	88
Education status (%)	
Illiterate	30.2
Below SSC	16.2
Only SSC	12.3
Graduate	19.5
Post graduate	16.8
Professional	5.0

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. Figure 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 4 or greater 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 ± test The alpha coefficient for the modified PHQ12 item was 0.863, which shows excellent reliability of this tool.

Factor Analysis

Factor analyses were conducted on the modified PHQ-12 items in order to examine any identifiable factor

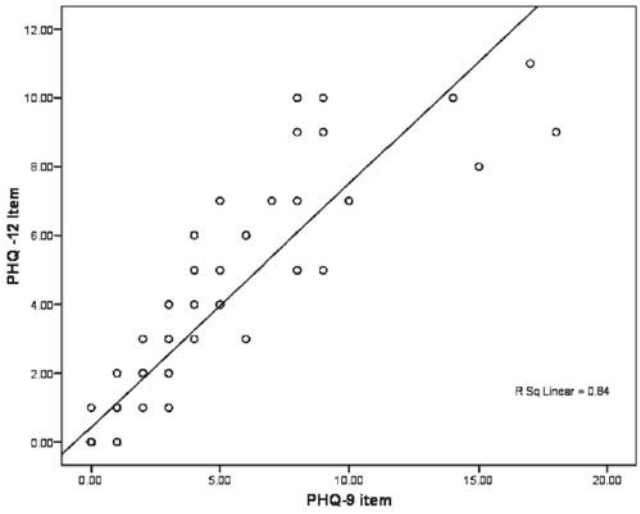
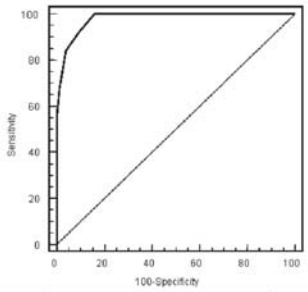


Fig. 1: Scatter plot –PHQ 12 items scores versus the PHQ-9 item scores.



Area under the ROC curve (AUC)- 0.979 ;
Standard Error- 0.020;
95% Confidence Interval- 0.929 to 0.997 ;
p value < 0.0001

Cut off Score	Sensitivity	Specificity	Positive predictive value	Negative predictive value
1	100	48	39.1	100
2	100	69.3	52.1	100
3	100	84	67.6	100
4	**92.2**	**90.7**	**76.7**	**97.1**
5	84	96.0	87.5	94.7
6	68	98.7	94.4	90.2
7	56	100	100	87.2
8	36	100	100	82.4
9	32	100	100	81.5
10	16	100	100	78.1

Fig. 2: Receiver operating characteristics (ROC) curve of the PHQ-12 items scale for detecting depression.

Table 3: Principle Component (Factor) Analysis of the Modified PHQ-12 item using cut off score ≥ 4

Symptom	Factor 1 (Depressive Symptoms)	Factor 2 (Physical Retardation)
Energy Loss	.79	
Worthlessness	.77	
Sadness	.74	
Lack of Concentration	.74	
Low Libido	.73	
Suicidal thoughts	.73	
Speech and Motor Retardation	.71	.33
Problems with Sleep	.69	.32
Physical Restlessness	.61	
Change in Appetite	.51	
Indulgence		.76
Withdrawal		.70

patterns and the results are presented in the Table 3. Principal component analysis resulted in 2 factors. The first factor yielded 10 items which explained 42% of total variance and the second factor yielded 2 items that explained 12% of the total variance. The first factor appeared to be mainly depressive symptoms (worthlessness, sadness, lack of concentration, low libido, suicidal thoughts and speech and motor retardation) whereas the second factor was focused exclusively on indulgence and withdrawal (overeating, weight gain and excessive sleep).

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,[13] in primary health care[18] and in medically ill patients.[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 4 or greater 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.[20] The use of the 2-item[14] and 15-item[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.[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[23]

in screening for depression. Using a cut-off value of 4 or greater, 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 Robin *et al*.[24] Indeed, a screening tool is considered reliable when the sensitivity range is 0.79 and the specificity > 0.63.[24]

Streiner *et al*,[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, anxiety) which help us to diagnose even mild depression in population based studies. Brevity coupled with its construct and criterion validity, makes the PHQ-12 item an attractive, dual purpose instrument for making diagnosis and assessing severity of depressive disorders in the general population.

The factor analysis data clearly indicates that the tool not only measures depression and its severity but also indicates the behaviour changes and hence the factorial validity of the questionnaire is well established. Moreover the PHQ-12 item scale demonstrated high psychometric similarity when compared with PHQ 9 item. Examination of the factor structure showed that the Modified PHQ 12 item scale is best considered and can be used as a unidimensional scale because the internal consistency is not compromised

There are however some limitations of this study i) there is a possibility that a proportion of participants might have underreported their depressive symptoms on both the PHQ-9 and the PHQ-12 interview because some subjects provided only vague descriptions of symptom durations or onset, while a few others gave very confusing and inconsistent estimates; ii) the tool may not be that useful for diagnosing depression in a clinic set-up or for use by psychiatric specialists because of the dichotomous response although this has to be separately assessed by another study. However the strength of the study is that, it was conducted as part of a large community based epidemiological study that is representative of the population of Chennai and therefore can be extrapolated to the rest of urban India. Also the entire study was carried out by non-medical people and still came out to be useful showing its simplicity and reliability for large scale use in a field setting outside of medical clinics and hospitals.

In conclusion, we have shown that the modified PHQ-12 is a valid and useful instrument in screening for depression amongst an urban south Indian population. A cut off score of 4 or more is recommended to identify people with depression. With increasing evidence suggesting that depression is increasing in the population, the PHQ-12, because of its validity, reliability, brevity and ease of administration will be a valuable tool for estimating depression in the community in India. Thus PHQ-12 can be used as a screening tool for large epidemiological studies and this is currently being done to assess prevalence of depression in Chennai in the CURES study.

Acknowledgement

We 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.Vetriselvei 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

1. Katon W, von Korff M, Lin E, Lipscomb P, Russo J, Wagner E. Distressed high utilizers of medical care: DSM-III-R diagnosis and treatment needs. *Gen Hosp Psychiatry* 1990;12:355-62.
2. Simon G, Ormel J, VonKorff M, Barlow W. Health care costs associated with depressive and anxiety disorders in primary care. *Am J Psychiatry* 1995;152:352-57.
3. Depression Guideline Panel. Depression in Primary Care: Volume 1, Detection and Diagnosis, Rockville, MD: US Dept of Health And Human Services, Public Health Service; Agency For Health Care Policy and Research;1993.Clinical Practice Guideline, No.5 AHCPR: Publication no.93-0550.
4. Spitzer RL, Willams JBW, Kroenke K, Spitzer RL, Williams JB, Kroenke K, *et al.* Utility of a new procedure care: the PRIME–MD 1000 study. *JAMA* 1994:272:1749-56.
5. Roness A, Mykletun A, Dahl AA. Help-seeking behaviour in patients with anxiety disorder and depression. *Acta Psychiatr Scand* 2005;111:51-8.
6. Wittchen HU, Lieb R, Wunderlich U, Schuster P. Comorbidity in primary care: presentation and consequences. *J Clin Psychiatry* 1999; 60:29-36.
7. Simon GE, Goldberg D, Tiemens BG, Ustun TB. Outcomes of recognized and unrecognized depression in an international primary care study. *Gen Hosp Psychiatry* 1999;21:97-105.
8. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
9. Beck AT, Steer RA, Brown GK. Manual for Beck Depression Inventory-II Psychological Corporation, San Antonio, TX, 1996.
10. Radloff LS. The CES-D scale: a self report depression scale for research in the general population. *Appl Psychol Measure* 1977;1:385-401.
11. Koenig HG, Meador KG, Cohen HJ, Blazer DG. Self-rated depression scales and screening for major depression in the older hospitalized patient with medical illness. *J Am Geriatr Soc* 1988;36:699-706.
12. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*

2001;16:606-13.

13. Martin A, Rief W Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *General Hospital Psychiatry* 2006;28:71–7.
14. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003;41:1284-92.
15. Lee PW, Schulberg HC, Raue PJ, Kroenke K. Concordance between the PHQ-9 and the HSCL-20 in depressed primary care patients. *J of Affective Disorders* 2007;99:139-45.
16. Deepa M, Pradeepa R, Rema M, Mohan A, Deepa R, Shanthirani S, *et al*. The Chennai Urban Rural Epidemiology Study (CURES) – Study Design and Methodology (Urban Component) (CURES-1). *J Assoc Physicians India* 2003;51:863-70.
17. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999;282:1737-44.
18. Wulsin L, Somoza E, Heck J. The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. Prim Care Companion. *J Clin Psychiatry* 2002;4:191-5.
19. Williams LS, Brizendine EJ, Plue L, Bakas T, Tu W, Hendrie H, *et al.* Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke* 2005;36:635-8
20. US Preventive Services Task Force, Screening for depression: recommendations and rationale. *Ann Intern Med* 2002;136:760-4.
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;64:258–66.
22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiving Operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
23. Holmes WC. A short, psychiatric, case-finding measure for HIV seropositive outpatients: performance characteristics of the 5-item mental health subscale of the SF-20 in a male, seropositive sample. *Med Care* 1998;36:237-43.
24. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, *et al.* The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* 1988;45:1069-77.
25. Streiner DL, Norman GR. Scaling responses. In: Health Measurement Scales: a practical guide to their development and use. 2nd ed. edition. Oxford, Oxford University Press; 1995:20-53.