

Comparison of the Personal Health Questionnaire and the Self Reporting Questionnaire in rural Pakistan

Nusrat Husain¹, Richard Gater², Barbara Tomenson², Francis Creed²

Clarke Division, Centre for Addiction and Mental Health, University of Toronto¹, Toronto, Ontario, Canada, University of Manchester², Division of Psychiatry and Behavioural Sciences, Manchester, England.

Abstract

Objective: To apply ROC analysis to select the best threshold scores for the PHQ and SRQ; to compare the sensitivity and specificity of the PHQ and SRQ against a criterion diagnosis of depressive disorder in a community sample in rural Pakistan, and to examine the influence of socio-demographic factors on misclassification.

Methods: The study used a two-stage design. Receiver Operating Characteristic (ROC) analysis was used to estimate the optimal threshold score and to compare the ability of the Self Reporting Questionnaire (SRQ) and the Personal Health Questionnaire (PHQ) to discriminate between cases of depressive disorder and non-cases.

Results: The results of the ROC analysis suggest that the SRQ is superior to the PHQ, and at the threshold of 5/6, the SRQ has superior sensitivity, negative predictive value and percentage agreement compared with the PHQ. When the SRQ threshold is raised it gains specificity, and at a cut-off threshold of 7/8 it is superior to the PHQ (5/6) in all validity coefficients and percentage agreement. Only gender and the presence of a confidant had a significant effect on misclassification using the SRQ among the cases. Both questionnaires performed better for females based on comparison of the areas under the ROC curves.

Conclusion: This study has demonstrated that the Urdu translations of both the PHQ and SRQ can be used as screening tests for depressive disorder in the Pakistani population. People with little or no education answer both somatic and psychological items with equal ease. In conclusion, the PHQ does not appear to have any advantage over the SRQ. (JPMA 56:366;2006).

Introduction

Depressive disorders in the developing world are a serious public health concern and are predicted to become the most common cause of disability by the year 2020¹. In developing countries with very limited resources, quick and inexpensive means of assessing depression are essential. Several different self-completed screening instruments have been developed. This study aimed to compare the performance of two screening questionnaires, the Personal Health Questionnaire (PHQ)² and the Self Reporting Questionnaire (SRQ)³, in the detection of depressive disorder. Since the PHQ was designed to detect depressive disorder, it is potentially more specific than SRQ, which was designed to detect a wider range of common mental disorders.

The PHQ is a 16-item self-report depression rating scale specifically designed for use in non-psychiatric clinics.² The items are based on the DSM-III-R criteria for depression, and subjects are asked to indicate (yes/no) symptoms they have experienced over the preceding two weeks. The PHQ has two parts; the first asks about the two key symptoms of sadness and loss of interest. The second part contains 14 questions about the other diagnostic

symptoms of depression and is only completed if at least one of the key symptoms is present. If one of the key symptoms is present and there are five or more other positive responses, the subject is likely to be currently depressed. In a primary care setting, the PHQ has a sensitivity of 83.9% and a specificity of 84.6% to detect major depressive disorder.⁴

The SRQ is one of the most widely used self-administered psychiatric questionnaires.³ The SRQ was developed as part of a collaborative study on strategies for extending mental health care coordinated by the WHO. It includes items indicative of non-psychotic mental disorders. It was developed to facilitate trans-cultural comparison by ensuring ease of translation and comparability of meaning between geographic regions.³ The SRQ has been compared with standardised assessments of psychiatric disorder in primary care and community studies in Pakistan and in other developing countries.⁵⁻⁸

The objectives of this study were: (i) to apply ROC analysis to select the best threshold scores for the PHQ and SRQ; (ii) to compare the sensitivity and specificity of the PHQ and SRQ against a criterion diagnosis of depressive

disorder in a community sample in rural Pakistan, and (iii) to examine the influence of socio-demographic factors on misclassification.

Methods

The study was a house-to-house survey of a geographically designated area of Mandra, 50 kilometres east of Islamabad. Urdu is widely understood and spoken, and local people speak Potohari, a dialect of Punjabi. Farming, military service, small local business ventures, employment in local cigarette factory and employment overseas are the main avenues of livelihood for majority of the population. The study was performed in 1996.

The sample was derived from the electoral list (age 18 years and above) held at the office of Union Council. Random numbers tables were used to select 300 names; half male, half female. A total of 267 were found to be currently living in the village. The remaining 33 people, mainly men, were working away semi-permanently, only returning every six months or so.

The study used a two-stage design. In the first stage, two research assistants, with local guides, introduced the study to potential participants, answered questions and gained verbal consent for inclusion in the study. The research assistants administered a social and demographic questionnaire along with Urdu translations of the PHQ and SRQ. The research assistants read out the questions to villagers who were illiterate.

In the second stage, PHQ high scorers (6 or more) were over-sampled in an otherwise random sample who were contacted for interview using the Psychiatric Assessment Schedule (PAS).⁹ It took 30-40 minutes to complete the PAS interview. PAS results were analyzed using the CATEGO computer program.¹⁰ The interviews were all performed by the first author who was blind to the screening questionnaire results. All interviews were conducted within 4 weeks of the first stage screening. Every effort was made to carry out the interview in privacy but in the case of female subjects a female research assistant accompanied the male interviewer. Psychosexual items in the interview had to be omitted in most female interviews. Cases of depressive disorder were defined as those with Index of Definition 5 or above using the PAS as the gold-standard criterion diagnosis throughout.

Receiver Operating Characteristic (ROC) analysis was used to estimate the optimal threshold score and to compare the ability of these two questionnaires to discriminate between cases of depressive disorder and non-cases. All calculations were carried out using expansion weights to adjust for the different proportions of screen positives and screen negatives that were interviewed using

the PAS.¹¹ Performance indices were calculated for both questionnaires using best cut-off determined by ROC analysis: sensitivity, specificity, positive predictive value, negative predictive value and percentage agreement. The effect of socio-demographic variables on misclassification by the questionnaires was first assessed by calculating the odds ratio of being a false negative among cases and of being false positive among non-cases using Fisher's exact test or logistic regression using expansion weights.

Results

Of 267 people approached, 258 (96.6%) completed the PHQ; 144 (56%) were female. A total of 107 were selected for second phase diagnostic interview, of whom 102 (95%) agreed; 29 males and 73 females. Table 1 shows the demographic details of the whole sample of 258 subjects, those interviewed at phase 1 only, and those interviewed at phase 2. No significant differences were found between the two sub samples, except that females were more likely to be interviewed at phase 2, because they were more likely to score highly on the PHQ screen.

ROC analysis was applied to the two screening tests (Figure). Sensitivity, specificity and percentage agreement between screen and CATEGO results were all adjusted, using expansion weights, to account for the different proportions of screen negatives and screen positives who were interviewed using the PAS. Unless this is done, screen negatives are under-represented in the second-phase sample, sensitivity is likely to be overestimated and specificity underestimated. Out of 101 high scorers (6 or more) on the PHQ, 57 were interviewed using the PAS. Thus the expansion weight for screen positive results should be $101/57 = 1.77$. Out of 157 low scorers, 45 were interviewed using the PAS, so weighting for screen negatives is $157/45 = 3.49$.

It is clear from Figure that the SRQ is a better screening instrument for these subjects than the PHQ because the curve is more concentrated in the top left hand corner of the graph area, which represents both high sensi-

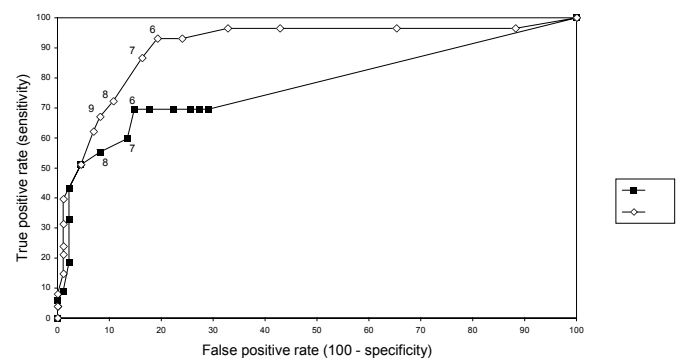


Figure. ROC curves using total SRQ or PHQ scores and sampling weights.

Table 1. Demographic details and comparison of subjects interviewed at the first phase only and those also interviewed in phase 2. Fisher's exact test p-value is shown for dichotomous variables.

	Whole sample n=258		1st phase only n=156		2nd phase n=102		Comparison of 1st and 2nd phase samples
	No	%	No	%	No	%	P
Male	114	44	85	55	29	28	<0.0005
Age 40 or more	109	42	72	46	37	36	0.12
Marital status ¹							
Single	31	12	22	14	9	9	
Married	207	81	124	80	83	81	0.25
Widowed/separated or divorced	19	7	9	6	10	10	
Educational level:							
Illiterate /can sign only	50	19	26	17	24	24	
Primary/can read Quran	82	32	46	29	36	35	0.12
at least secondary	126	49	84	54	42	41	
Unemployed	23	9	15	10	8	8	0.66
4 or more children ²	153	62	89	60	64	66	0.42
Lost a child	42	16	25	16	17	17	1.0
Separation between parents before age 17 ³	27	11	20	14	7	7	0.14
Separation from parents before age 17	93	36	60	39	33	32	0.35
Subject has a confidant	164	64	95	61	69	68	0.29
Monthly income ⁴							
< 1000 rupees	21	8	12	8	9	9	
1000 -5000 rupees	187	75	111	74	76	78	0.54
> 5000 rupees	41	17	28	18	13	13	
Living situation: 3 or more people per room	80	31	43	28	37	36	1.0

1 1 subject in the 1st phase only group did not supply information on marital status.

2 8 subjects in the 1st phase only group and 5 in the 2nd phase group did not supply information about number of children.

3 13 subjects in the 1st phase only group and 5 in the 2nd phase group did not supply information about separation between parents.

4 5 subjects in the 1st phase only group and 4 in the 2nd phase group did not supply information about monthly income.

Table 2. Validity and agreement coefficients for PHQ and SRQ against the PAS criterion diagnosis of depressive disorder.

Threshold	Specificity	Sensitivity	PPV	NPV	Percentage agreement
PHQ:					
5/6	85.2	69.6	78.9	77.8	78.2
6/7	86.6	59.8	78.8	72.0	74.4
7/8	91.8	55.3	84.8	71.4	75.4
8/9	95.5	51.1	90.2	70.5	75.5
SRQ:					
5/6	80.8	93.1	77.9	94.1	86.0
6/7	83.7	86.6	79.7	89.5	85.0
7/8	89.3	72.2	83.6	80.9	81.9
8/9	91.8	67.0	86.3	78.4	81.1

PPV = positive predictive value

NPV = negative predictive value

corner of the graph area, which represents both high sensitivity and high specificity. The best threshold can be identified from the ROC curve by selecting the threshold

that is the shortest distance from the top left hand corner of the plot area (i.e. coordinates 0,100). Using the SRQ, the threshold of 5/6 appears to be good, although 6/7, 7/8 or 8/9 may also be worth considering, depending on whether it is regarded as important to be sensitive or specific. These have been labeled on Figure. Choosing between different thresholds may be influenced by the characteristics required of the screening test: selection of a lower threshold would increase sensitivity with some loss of specificity, whereas raising the threshold increases specificity with loss of sensitivity. For the PHQ, the cut-off of 5/6 actually used in the first-phase screen in this study seems a good choice.

Validity coefficients for various thresholds of each questionnaire are shown in Table 2. At a threshold of 5/6 the SRQ has superior sensitivity, negative predictive value and percentage agreement compared with the PHQ (5/6). At a threshold of 8/9 the SRQ has nearly as good sensitivity as the PHQ (threshold 5/6), but much better specificity.

Of the 102 subjects, 22 were misclassified using the PHQ with a cut-off of 5/6 (10 false negatives and 12 false

positives). Using the SRQ (5/6), 17 were misclassified (2 false negatives and 15 false positives) and using SRQ (8/9), 18 were misclassified (11 false negatives and 7 false positives). Among the socio-demographic variables investigated, only gender and the presence of a confidant had a significant effect on misclassification using the SRQ among the cases. There was no difference in misclassification by gender using the PHQ. Age, marital status, overcrowding, education, income, employment status, having four or more children, loss or separation from parents before the age of 17 years or loss of a child were not associated with differences in misclassification for either screening test using logistic regression with gender-specific expansion weights.

As the SRQ performed differently according to the gender of the respondent, the ROC curves of both questionnaires were plotted separately for males and females. Both questionnaires performed better for females than males, based on comparison of the areas under the ROC curves. Using the PHQ, the cut-off of 5/6 would be suitable for both males and females. In contrast, the best threshold using the SRQ appears to be different for males and females. While 8/9 would be a good threshold for females, it would be very poor for males due to lack of sensitivity (15.8%). A much lower threshold, such as 5/6 or even 3/4 would appear to be better for males.

Discussion

This study has demonstrated that the Urdu translations of both the PHQ and SRQ can be used as screening tests for depressive disorder in the rural Pakistani population. Both scales were found to be very acceptable to the participants who had no difficulty in understanding its purpose and completing it. People with little or no education answer both somatic and psychological items with equal ease.

The results of the ROC analysis suggest that the SRQ is superior to the PHQ, and at the threshold of 5/6, the SRQ has superior sensitivity, negative predictive value and percentage agreement compared with the PHQ. When the SRQ threshold is raised it gains specificity, and at a cut-off threshold of 7/8 it is superior to the PHQ (5/6) in all validity coefficients and percentage agreement.

In addition to the ability of a screening questionnaire to identify a probable case, the factors that influence misclassification also need to be considered when choosing between questionnaires. Sociodemographic variables have been associated with the misclassification of subjects by psychiatric questionnaires.¹²⁻¹⁴ Studies using the General Health Questionnaire in Europe have shown that its sensitivity is lower in men than in women¹⁵⁻¹⁷ and in Brazil poorly educated people are more likely to be misclassified

as false positive and women as false negative.¹² In this study Both PHQ and SRQ perform better with females than males. For the PHQ the cut-off 5/6 was the best for both genders, but it may be better to choose a lower threshold for the SRQ among men. Within the limits of the modest sample size, neither the SRQ nor PHQ varied in performance with respect to other factors associated with prevalence of depression, including age, marital status, overcrowding, education, income, employment status, having four or more children, loss or separation from parents before the age of 17 years or loss of a child.

In conclusion, the PHQ does not appear to have any advantage over the SRQ. Similarly, it was reported that the Bradford Somatic Inventory did not have any advantage over the SRQ, which has now been widely used in developing countries.^{18,19} This study has validated the SRQ in a new setting in Pakistan, but in the light of the sample size, further research on the use of screening instruments in such settings is required.

References

1. WHO, World Bank "The Global Burden of Disease". Edited by Murray CJL and Lopez AD. Published by the Harvard School of Public Health on behalf of the World Health Organisation and the World Bank, 1996.
2. Simpson N. Validation of a New Self-rating Questionnaire to Detect Psychiatric Illness in General Practice: The Personal Health questionnaire. Msc.thesis; University of Manchester, 1984.
3. Harding TW, De Arango, MV, Baltazar, J et al. Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries, 1980, *Psychological Medicine*, 10, 231-241.
4. Husain, N, Creed, F, Tomenson, B. Adverse social circumstances and depression in UK persons of Pakistani origin. *Br J Psychiatr* 1997, 171:434-433.
5. World Health Organization. A User's Guide to the Self Reporting Questionnaire (SRQ). Divisional of Mental Health, WHO Geneva, 1994.
6. Minhas FA, Iqbal K, Mubbashar MH. Validation of Self-Rating Questionnaire in primary care settings of Pakistan. *Pak J Clin Psychiatr*, 1995, 5,60-69.
7. Saeed K, Gater R, Hussain A, Mubbashar M. The prevalence, classification and treatment of mental disorders among attenders of native faith healers in rural Pakistan. *Social Psychiatry and Psychiatric Epidemiology*; 2000, 35:480-485.
8. Mumford DB, Minhas FA, Akhtar I, Akhtar S, Mubbashar MH. Stress and psychiatric disorder in Urban Rawalpindi. *Community Survey. Br J Psychiatr* 2000,177:557-62.
9. Dean C, Surtees PG., Sashidharan SP. Comparison of research diagnostic systems in an Edinburgh community sample. *British Journal of Psychiatry*; 1983,142, 247-56.
10. Wing JK, Cooper JE, Sartorius N. The description and classification of psychiatric symptoms: An introduction manual for the PSE and CATEGO System. London: Cambridge University Press. 1974.
11. Bisoffi G, Mazzi MA, Dunn G. Evaluating screening questionnaires using Receiver Operating Characteristic (ROC) curves from two-phase (double) samples. *International Journal of Methods in Psychiatric Research*, 9:121-133, 2000.
12. Mari JJ, Williams P, A validity study of a psychiatric screening questionnaire (SRQ-20) in primary care in the city of Sao Paulo. *Br J Psychiatr* 1986,148:23-6.
13. Goldberg D, Williams P. A User's Guide to the GHQ. NFER-Nelson, Windsor, 1988.
14. Dohrenwend BP. The problem of validity in field studies of psychological disorders revisited. *Psychol Med* 1990, 20:195-208.

15. Vazquez-Barquero JL, Diez-Manrique JF, Pena C, Quintanal RG, Labrador Lopez M; Two stage design in a community survey. *Br J Psychiatr* 1986;149:88-97.
16. Hobbs P, Ballinger CB, Smith DMW. Factor analysis and validation of the General Health Questionnaire in women and general practice surveys. *Br J Psychiatr* 1983;142, 257-64.
17. Hobbs P, Ballinger CV, Greenwood C, Martin B, McClure A. Factor analysis and validation of the General Health Questionnaire in men: a general practice survey. *Br J Psychiatr* 1984;144, 270-5.
18. Mumford DB, Saeed K, Ahmad I. Stress and psychiatric disorder in rural Punjab. A community survey. *Br J Psychiatr* 1997 170:473-78.
19. Mumford DB, Bavington JT, Bhatnagar KS, Hussain Y, Mirza S, Naraghi.M. The Bradford Somatic Inventory. A multi-ethnic inventory of somatic symptoms reported by anxious and depressed patients in Britain and the Indo-Pakistan subcontinent, 1991. *Br J Psychiatr* 158:379-86.