

## Original Article

# Validation of the Sinhala version of the 12-item General Health Questionnaire

H T C S Abeysena<sup>1</sup> P Jayawardana<sup>1</sup> U Peiris<sup>2</sup> A Rodrigo<sup>2</sup>

<sup>1</sup>Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, <sup>2</sup>Department of Psychiatry, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

**Keywords:** General Health Questionnaire; mental disorders; primary care; stress; validity

## Abstract

### Objective

To determine the criterion validity of the Sinhala version of the General Health Questionnaire-12 (GHQ-12).

### Methods

This was a descriptive cross-sectional study including 374 patients (between the ages 18 to 75 years) attending the outpatient department of the Colombo North Teaching Hospital during the period June 2009 to September 2010. The Sinhala version of the GHQ-12 was completed by the participants. The 'Clinical Examination' was done by the designated psychiatrist (blindly to the GHQ score) by employing the ICD 10 classification of 'Depression/Anxiety and Social Dysfunction' as a reference standard. Cut-off values for the GHQ-12 were determined by applying the Receiver Operation Characteristic (ROC) curves.

### Results

The optimal cut-off level for GHQ-12 was 1/2 with a sensitivity of 74% (95%CI 66%-81.5%), and a specificity of 71% (95%CI: 65.4%-76.6%). Optimal cut-off levels were not changed based on gender, age and educational levels. The area under the ROC curve for GHQ-12 was 0.79 (95% CI: 74%-84%). Despite the above mentioned statistics/measurements, GHQ scores with specific likelihood ratios of >1 were obtained only for the total scores of  $\geq 3$ .

### Conclusions

The GHQ-12 displays its efficacy when used among the Sinhala speaking Sri Lankan primary care settings for assessing psychiatric disorders.

Corresponding Author: HTCS Abeysena, E-mail<[chrishanthaabeysena@yahoo.com](mailto:chrishanthaabeysena@yahoo.com)>

Received: Feb. 2014, Accepted revised version: Dec. 2014, Published: December 2014

Competing Interests: Authors have declared that no competing interests exist

## Introduction

The General Health Questionnaire (GHQ) is a self-administered questionnaire that is widely used to detect potential non-psychotic psychiatric disorders. Although the original GHQ questionnaire contained 60-items, 30, 28- and 12-item versions have been derived from it subsequently. These versions have been used globally in diverse settings.

The Sinhala translation of the GHQ-30 has been widely used for research purposes in Sri Lanka. Nevertheless, GHQ-12 has several advantages over the GHQ-30, for example, the former is brief, simple and less time consuming to administer. These features make the use of GHQ-12 more attractive for both research purposes and screening in busy clinical settings for psychological morbidity

The 12 items included in the GHQ-12 have been extracted from the 30 item version and validated to screen for the presence of psychological morbidity. The translation of screening tool terms into the local language requires retention of the conceptual meaning while adapting

the term to the relevant culture. In trying to achieving this, even the slightest change in the format can alter the comprehension of the concept and, thus affect the accuracy of the responses obtained. Therefore, it is important to validate all translated terms of research instruments, so that the questionnaire may be accurately applied and interpreted.

Validity is the extent to which a test measures what it claims to measure. Criterion validity refers to the accuracy of a tool when applied against a reference standard<sup>1</sup>. With regards to the GHQ, validity indicates the extent to which the test scores accurately estimate an individual's current psychological state in relation to the criterion. There are four possible methods of scoring the GHQ. The authors advocate GHQ scoring method (0-0-1-1) and the suggested cut-off value was 1/2. The objective of this study was to determine the criterion validity of the Sinhala version of the GHQ-12.

## Methods

A descriptive cross sectional study was conducted at the Out Patient Department (OPD) of the Colombo North Teaching Hospital (CNTH), Ragama, during the period June 2009 to September 2010. The study participants included both males and females between the ages of 18 and 75 years, who were able to read and understand the Sinhala language.

The sample size was computed using the formula for estimating a population proportion of the health-related event of interest (psychological morbidity) with absolute precision<sup>2</sup>. Psychological morbidity rate for Sri Lanka was considered to be 30%, and, based on that, the computed sample size was 323 for a confidence level of 95% and an absolute precision of 5%.

All patients (who consented to be in the study) were registered at the CNTH of Ragama for OPD visits during the study period, and those eligible for inclusion, were recruited using the convenience sampling method until the required sample size was reached. During this process around 10 to 15 patients/ day were recruited. Sinhala version of the GHQ-12, which is a self-administered questionnaire, was rated on a four-point Likert. In addition socio-demographic data were also collected. A record sheet was used to document the clinical assessments made by the relevant psychiatrist with regards to the presence of depression/anxiety/social dysfunction. This was considered as the reference standard and the diagnostic criteria were based on the ICD 10 classification. The psychiatrist was blinded to the results of the GHQ score.

Statistical analysis was performed using the statistical package 'SPSS Windows Version 16'. The Likert scale was transformed into the GHQ score (0, 0, 1, 1) for this analysis. Sensitivity and specificity of the instrument with 95% confidence intervals (CI) were determined for different cut-off values. However, the optimal cut-off value for the instrument was determined using the Receiver Operation Characteristic (ROC) curves, which were based on the trade-off between sensitivity and 1- specificity. The area under the curve (AUC), which represents a summary measure of the ability of the GHQ to discriminate between cases and non-cases, was the basis applied to determine the accuracy of the curve. The lowest distance between the point (0, 1) and any point on the ROC curve was the criteria for determining optimal cut-off value. Sex, educational and age group differences also were explored.

Written informed consent was obtained from all the research participants and confidentiality of data obtained were achieved by concealing their identity, using a coding system. Ethical clearance was obtained from the Ethics Review Committee, Faculty of Medicine, University of

Kelaniya. The psychiatrist in attendance took necessary steps to attend to those participants who required treatment.

## Results

The total number of OPD patients who were invited to the study during the study period was 432. Of them, 22 patients were not willing to participate, which gives a non-participant rate of 5.0 %. Therefore, the total patients recruited to the study was 410, out of which, 12 (2.9%) were not accessible for psychological assessment by the nominated psychiatrist. Of the 410 patients, 24 (6.1%) did not have the necessary data for one or more items of the GHQ. Hence, these 36 (9.0%) were excluded from the final analysis, which left a study sample of 374.

The median age of the study sample was 32.5 years (IQR = 21years). The highest proportion (28.5%; n = 117) were in the age group 21 to 30 years. Fifty nine percent (n = 242) of the study sample were Buddhists and 26.6% (n = 110) Catholics or Christians. Two hundred and twenty two (54 %) had studied up to the General Certificate Examination (Ordinary Level) and 242 (65%) were females. The median GHQ score was 1 (IQR = 4).

Of the 374 patients who were assessed by psychiatrists 83 (22%) had at least one psychological morbidity issue, while 27 (7.2%) had two and 16 (4.3%) had three such issues, respectively. Among the psychiatric disorders identified, 118 (31.6%) suffered from depression, 41 (11%) from anxiety and 26 (7%) from a social dysfunction. Of the 374 participants 126 (33.6%) had psychological morbidities according to the reference standard. The median GHQ score was one (IQR 4) and 130 (35%) and 43 (11.5%) participants scored zero and above six respectively.

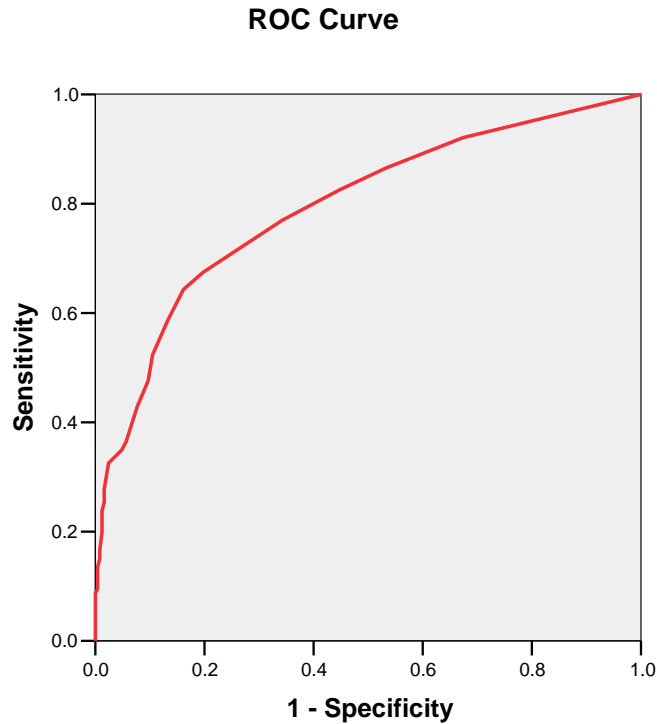
## Determination of the optimal cut-off level for GHQ-12

The GHQ-12 score was categorised into two groups at various cut-off points. The corresponding sensitivity, specificity and likelihood ratios are shown in Table 1.

**Table 1: Distribution of psychological morbidity by different cut-off values of GHQ- 12, sensitivity, specificity and likelihood ratios**

GHQ Score	Reference Standard		Sensitivity %	Specificity %	Misclassification Rate	Likelihood Ratio
	Yes (n=126) n (%)	No (n=248) n (%)				
≥1	111 (88.1)	133 (53.6)	88.0	46.4	32.7	1.64
≥2	93 (73.8)	72 (29.0)	73.8	71.0	27.6	2.54
≥3	81 (64.3)	45 (18.1)	64.3	82.0	27.0	3.55
≥4	72 (57.1)	33 (13.3)	57.1	86.7	28.0	4.29
≥5	52 (41.3)	19 (7.7)	41.3	92.3	33.2	5.36
≥6	46 (36.5)	19 (7.7)	36.5	95.6	34.0	8.29

The AUC of the ROC curve was 79% (95% CI: 74% –84%) (Figure 1).



**Figure 1: ROC Curve**

The optimal cut-off level for the GHQ-12 was 1/2, which gave a sensitivity of 74% (95% CI 66% -81.5%) and a specificity of 71% (95% CI: 65.4% – 76.6%). The GHQ-12 misclassified 27.6% of the subjects either as false positives or false negatives at this cut off level (Table 1). The corresponding likelihood ratio for positive tests was 2.5 (95% CI: 2.0 – 3.2). However, the lowest misclassification rate was observed when the cut-off level was 2/3 and, at which, the sensitivity, specificity and likelihood ratios for positive tests were 64% (95% CI: 56.4%–73%), 82% (95% CI: 77%–87%) and 3.6 (95% CI 2.6– 4.8), respectively.

The AUCs were above 70% for the different subgroups analysed according to selected socio-demographic variables (Table 2). The best cut off level for all subgroups was 1/2 except for males and Catholics/Christians.

**Table 2: Optimal cut-off values, sensitivity, specificity and area under the curve (AUC) of the Sinhala version GHQ 12 by different levels of socio-demographic variables**

Variables	cut off	Sensitivity %	Specificity %	Area Under the Curve %	95% Confidence Interval
<b>Sex</b>					
Male (n=127)	2/3	73.0	76.0	80.0	0.71-0.88
Female (n=242)	1/2	70.0	73.0	79.0	0.72-0.85

<b>Age (Years)</b>					
<32 (n = 192)	1/2	76.0	71.0	80.0	0.73-0.87
≥ 32 (n= 182)	1/2	71.0	71.0	78.0	0.71-0.85
<b>Educational Level</b>					
<grade 10 (n = 206)	1/2	80.0	76.0	71.0	0.73-0.86
≥grade 10 (n = 155)	1/2	71.0	73.0	77.0	0.68-0.87
<b>Income</b>					
Low (n = 99)	1/2	75.0	71.0	79.0	0.70-0.88
High 10 (n = 101)	1/2	88.0	71.0	88.0	0.80-0.96
<b>Religion</b>					
Buddhists (n = 233)	1/2	79.0	76.0	83.0	0.77-0.89
Catholics (n = 90)	3/2	63.0	78.0	72.0	0.61-0.83

Table 3 presents the GHQ-12 score specific (multilevel) to likelihood ratio positive values, which was less than one for up to a GHQ-12 score of two.

**Table 3: Multilevel likelihood ratios for GHQ-12**

<b>GHQ Score</b>	<b>Reference Standard</b>		<b>Likelihood Ratio</b>
	<b>Yes (n=126) n (%)</b>	<b>No (n=248) n (%)</b>	
0	15 (11.9)	115 (46.4)	0.26
1	18 (14.3)	61 (24.6)	0.58
2	12 (9.5)	27 (10.9)	0.87
3	9 (7.1)	12 (4.8)	1.48
4-6	34 (26.9)	28 (11.3)	2.39
7-12	38 (30.1)	05 (2.0)	15.0

## Discussion

According to ROC analysis, the best cut off level was 1/2 of the GHQ-12, with a sensitivity of 74%, a specificity of 71% and a misclassification rate of 27.6%. Regarding all the subgroups considered, the best cut off level remained at 1/2 except for males and Catholics/Christians. This is consistent with the recommendations of the authors of the original GHQ. Most other studies too reported the same cut off value, although some studies have gone even up to values of 5/6. An Australian study<sup>3</sup> reported a cut-off level of 0/1, with a sensitivity of 75% and a specificity of 70%, in a representative community of Australians. According to a WHO study<sup>4</sup> thresholds varied in different centres from 1/2 (Shanghai and Nagasaki) to 6/7 (Bangalore). In most centres the threshold level was 1/2, while in others it was 2/3 or 3/4. However, the

recommended value determined by a later study<sup>4</sup> was also 1/2 with a sensitivity of 83.5% and a specificity of 75%.

A study<sup>5</sup> conducted among ethnic Indian women living in the United Kingdom reported a threshold level of 2/3 with a sensitivity of 96.7% and a specificity of 90%. These values are higher than those reported in our study and this difference may be due to varying definition of the study population and sampling methods. The above study employed a test-based sampling method. A study conducted in South India<sup>6</sup> based on the Tamil version of the GHQ-12, also reported an optimal cut off of 2/3 with a sensitivity of 87% and a specificity of 79%. The similarities observed between the above two studies may be attributed to the cultural similarities of the participants (both study populations were of Indian origin). A study<sup>7</sup> conducted on a Polish working population also reported a cut-off value of 2/3 with a sensitivity of 64% and a specificity of 79%. The corresponding figures for our study (at a cut off level of 2/3) were 64% (sensitivity) and 82% (specificity). Furthermore, the lowest misclassification rate in our study was also observed at a cut off level of 2/3.

We also found that the GHQ score specific to likelihood ratio was less than one for up to a GHQ score of two and then started gaining momentum, giving values higher than one over a score of  $\geq 3$ . These results are consistent with what was reported by Furukawa<sup>8</sup> in the WHO study. Multilevel likelihood ratios, however, give more information and are an alternative to the fixed threshold approach. A study conducted in Australia<sup>9</sup>, including a community sample, reported a stratum-specific GHQ score of likelihood ratios for people with physical co-morbidity who had been seeking help from a doctor within the past four weeks. These results were consistent with that of our study. Our study sampled OPD patients in the primary care setting, which is comparable to the study sample of the above study.

Although the recruitment of our study participants were not based on a probability sampling method, the non-response rate observed was minimal, which is likely to negate the selection bias that would otherwise, affect the external validity of the study. Furthermore, as the psychiatric assessment was carried out blind to the GHQ status, one could assume that the threat to the internal validity of the study would have been minimal. The non-purposive selection of the sample, however, would have exerted negligible spectrum bias. All of the above would have contributed to improving the generalisability of the results for a primary care setting.

The estimate of prevalence of psychiatric disorders is dependent on the criteria for 'casesness' that is used for validation. In our study, the reference standard was an assessment made by a psychiatrist using consistent standard criteria. The various methods give different results because of diversity in defining 'casesness'<sup>10</sup>. Goldberg<sup>4</sup> reported that the 'Clinical Interview Schedule' is almost identical to that which is used for the 'Present State Examination' in determining the validity. We agree with Goldberg et al<sup>10</sup> that the GHQ can be used efficiently in developing countries, and that sex, age and educational level do not affect the validity of the GHQ.

Therefore, considering the recommendations made in the User Guide of the General Health Questionnaire<sup>11</sup> we give priority to sensitivity in preference to specificity and set the threshold level at 1/2 of the GHQ-12, for purposes of case detection. Alternatively, the threshold level of 2/3 may be used. In conclusion, the GHQ-12 displays adequate validity (when used among Sinhala speaking Sri Lankan primary care setting attendees) for screening psychiatric disorders.

## Acknowledgements

This study was funded by the University of Kelaniya, Sri Lanka. We are grateful to the Research and Publication Committee and the research assistants of the University of Kelaniya, Sri Lanka.

## References

1. Abramson JH, Abramson ZH. Survey methods in community medicine. London: Churchill Livingstone; 1990.
2. Lwanga SK, Lemeshow S. Sample size determination in health studies: A practical manual. Geneva: World Health Organization; 1991.
3. Donath S. The validity of the 12-item General Health Questionnaire in Australia: a comparison between three scoring methods. *Aust N Z J Psychiatry* 2001;35:231-35.  
Doi: <http://dx.doi.org/10.1046/j.1440-1614.2001.00869.x>
4. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M et al. Validity of two versions of the GHQ in the WHO study of Mental illness in general health care. *Psycho Med*1997; 27:191-97.  
doi:[dhttp://dx.doi.org/10.1017/S0033291796004242](http://dx.doi.org/10.1017/S0033291796004242)
5. Jakob KS, Bhugra D, Mann H. The validation of the 12-item General Health Questionnaire among ethnic Indian women living in the United Kingdom. *Psychol Med* 1997;27:1215-17
6. Kuruvilla A, Pothan M, Philip K, Braganza D, Joseph A, Jacob KS. The validation of the Tamil version of the 12 item General Health Questionnaire. *Indian Journal of Psychiatry* 1999; **41**(3):217-21
7. Makowska Z, Merecz D, Mościcka A, Kolasa W. The validity of general health questionnaires, GHQ-12 and GHQ-28, in mental health studies of working people. *Int J Occup Med Environ Health* 2002;**15**(4):353-62
8. Farukawa T, Goldeberg DP. Cultural invariance of likelihood ratios for the general health questionnaire. *Lancet* 1999;353:561-62.  
doi:[http://dx.doi.org/10.1016/S0140-6736\(98\)05470-1](http://dx.doi.org/10.1016/S0140-6736(98)05470-1)
9. Furukawa TA, Andrews G, Goldberg DP. Stratum-specific likelihood ratios of the General Health Questionnaire in the community: help-seeking and physical co-morbidity affect the test characteristics. *Psychol Med* 2002;32:743-48  
doi:<http://dx.doi.org/10.1017/S0033291702005494>
10. Goldberg DP, Oldhinkel T, Ormel J. Why GHQ threshold varies from one place to another. *Psychol Med* 1998;28:915-21.  
doi:<http://dx.doi.org/10.1017/S0033291798006874>
11. Goldberg D, Williams P. A user's guide to the General Health Questionnaire, Windsor: NFER-Nelson;1991

