ORIGINAL ARTICLE

THE SENSITIVITY, SPECIFICITY AND RELIABILITY OF THE MALAY VERSION 12-ITEMS GENERAL HEALTH QUESTIONNAIRE (GHQ-12) IN DETECTING DISTRESSED MEDICAL STUDENTS.

Muhamad Saiful Bahri Yusoff*, Ahmad Fuad Abdul Rahim*, Mohd Jamil Yaacob**

***Medical Education Department of School of Medical Sciences 16150 Kubang Kerian, Kelantan, Malaysia.**Department of Psychiatry, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia.**

Abstract

Objective: **To determine the sensitivity, specificity and internal consistency of the Malay version GHQ-12 among medical student population. This study determined the appropriate GHQ-12 score to detect distressed medical students.** ***Methods*****: The Malay version of GHQ-12 was derived based on two sources which were the original version GHQ-12 and the validated Malay version 30-items GHQ. The GHQ-12 and the Malay version Beck Depression Inventory-II (BDI-II) were administered to a total of 141 medical students. Distress diagnoses were made based on the Malay version BDI-II. ROC curve analysis was applied to determine the sensitivity and specificity of the GHQ-12 by testing against the BDI-II. Reliability analysis (Cronbach's alpha and item total correlation) was applied to test internal consistency of the GHQ-12. The analysis was done using SPSS version 12.*** ***Results*****: The GHQ-12 sensitivity and specificity at cut-off point of 3/4 was 81.3% and 75.3% respectively with positive predictive value (PPV) of 62.9% as well as area under ROC curve more than 0.7. The Cronbach's alpha value of the GHQ-12 was 0.85.*** ***Conclusion*****: This study showed the Malay version GHQ-12 is a valid and reliable screening tool in detecting distressed medical students. The GHQ-12 score equal to or more than 4 was considered as significant distress.** ***ASEAN Journal of Psychiatry, Vol.11 (1): Jan – June 2009: XX XX.***

Keywords: **Reliability,validity,medical students,General Health Questionnaire (GHQ)**

Introduction

The General Health Questionnaire (GHQ) is widely used internationally and locally [1-7] to measure mental health status especially in detection of emotional disorders such as distress. Since Goldberg introduced the GHQ in 1978, it has been translated into 38 different languages, testament to the validity and reliability of the questionnaire [4]. Reliability coefficients of the questionnaire have ranged from 0.78

to 0.95 in various studies [4]. It has four versions based on the number of items; GHQ-60, GHQ-30, GHQ-28 and the shortest version GHQ-12. Each item is accompanied by four responses, typically being 'not at all', 'no more than usual', 'rather more than usual' and 'much more than usual'. There are two recommended methods for scoring the GHQ. The first scoring method ranged from 0 to 3 respectively. The second scoring method was binary scoring method (with the two least symptomatic answers scoring 0 and the two most symptomatic answers scoring 1 – i.e. 0-0-1-1). The total possible score on GHQ-28 ranges from 0 to 84 and allows for means and distributions to be calculated, both for the global total, as well as for the four sub-scales (somatic symptoms, anxiety/insomnia, social dysfunction and severe depression). Using the alternative binary scoring methods the 28- and 30-items versions classify any score exceeding the threshold of 4 as achieving 'caseness'. The caseness threshold is 3 for the 12-item version. The shortened version work was found to be as reliable as the long version in detecting distress [3]. The GHQ-12 was commonly used as its validity is well-established internationally [8, 9] and locally [1, 2]. It was also used because of its popular use in student sample [9] and young populations in the community [8]. Furthermore it is simple, easy to understand, short and straightforward to complete. However, most of the validation studies have been done in the western countries [7] and hardly found in the Malaysian studies especially for medical student population.

The purpose of this study is to determine the sensitivity, specificity and reliability of the Malay version GHQ-12 among medical student population particularly in Malaysia.
It is also to determine the appropriate GHQ score in detecting distress in the population.

Methods

Development of the Malay version GHQ-12

The Malay version of GHQ-12 was derived based on two sources which were the original version GHQ-12 [10] and the validated Malay version 30-items GHQ [1]. These sources have provided the blueprint for the development of the Malay version GHQ-12. The language of item was made simple and suitable to express the concept implied.

Expert evaluation of the items

In order to establish the content validity of the Malay version GHQ-12, the items were subjected to jury technique. The experts were drawn from the field of Psychiatry. The jury group was formed by a psychiatrist and only psychiatrists who have more than 3 years experience were chosen as jury members. The items of the GHQ-12 were rated under 4 categories of responses; *tiada langsung* (not at all) , *tidak lebih dari biasa* (no more than usual), *lebih dari biasa* (more than usual), *sangat lebih dari biasa* (much more than usual) for statements: 1, 2, 7, 10, 11 and 12 as shown in Table 2, whereas, for the rest of the statements the responses were *lebih dari biasa* (more than usual), *tidak lebih dari biasa* (no more than usual), *kurang dari biasa* (less than usual) and *sangat kurang dari biasa* (much less than usual). The scoring method was a binary scoring method whereby the two least symptomatic answers score 0 and the two most symptomatic answers score 1 – i.e. 0-0-1-1. The minimum GHQ-12 total score was 0 and the maximum GHQ-12 total score was 12.

Preliminary try-out

The items were arranged as shown in Table 2, and administered to a sample of 48 medical students to check the comprehensibility and face validity during face-to-face session. The students were encouraged to express their doubts freely. The students were given 30 minutes to go through all the items individually and another 30 minutes to express their thoughts about the items to the investigator. Necessary modifications were made with the experience gained through this preliminary try-out.

Validation study

The Malay version GHQ-12 used for this study contained 12 items. Population of this study was 1065 medical students of 2008/2009 academic session in the School of Medical Sciences, Universiti Sains Malaysia. Proper instructions were given before the administration of the scale. The subjects were asked to respond to all the statements and no time limit was imposed. During the time of administration the investigator gave proper assistance and directions whenever and wherever necessary.

Sample size

Sample size calculated based on recommended ratio which was 10 subjects per item (11) with 20 percent dropout rate was 150 subjects. Convenient sampling method was applied; 50 second year medical students and 100 fifth year medical students were asked to participate in this study.

Collection of data

The investigator requested 150 medical students from second year and fifth year to fill in the GHQ-12. Completion of the questionnaire was voluntary and would not affect the students' progress in the course. A face-to-face session was held with the students in a hall. Data was collected by guided self-administered questionnaire. The time taken by the students to fill in the questionnaire was around 10 minutes. The questionnaires were collected on the same day. Verbal consent was taken from the students. The investigator obtained permission and clearance from the School of Medical Sciences and Human Ethical Committee of Universiti Sains Malaysia.

Reliability analysis

Reliability analysis was done using SPSS version 12 to determine the internal consistency of the items measured by using Cronbach's alpha coefficient. For an estimation of reliability, statistical reliability of individual items was done. Items with Cronbach's alpha value if item-deleted could determine which statement highly contributed to the alpha value. If the Cronbach's alpha value for those items-deleted were decreased, it indicated that the items were highly contributed to alpha value. In contrast, if the Cronbach's alpha value for those items-deleted increased, it indicated that the items poorly contributed to alpha value. The items of the GHQ-12 were considered to represent a measure of high internal consistency if the Cronbach's alpha value was more than 0.7 [12].

Sensitivity and specificity analysis

Distress detection was made based on the Malay version BDI-II. The BDI-II score equal to and more than 9 was considered as positive score for significant distress [13]. In order to determine the sensitivity and specificity, the GHQ-12 was tested against the distress diagnoses made by the BDI-II. The Receiver Operating Characteristics (ROC) curve analysis was done using SPSS version 12 to determine the sensitivity,

specificity, and area under ROC curve. The sensitivity, specificity and area under ROC curve value more than 0.70 was considered as having an acceptable predictive and discriminative value (14). The negative and positive predictive values were calculated manually via Microsoft Excel software.

Results

A total of 141 (94%) out of 150 medical students responded. 99 (70.2%) were female students. 92 (65.2%) were fifth year medical students and 49 (34.8%) were second year medical students. 86 (61.0%) were Malay, 53 (37.6%) were Chinese, 2 (1.4%) were Indian and 3 (0.3%) were others. 86 (61.0%) were Muslim, 37 (26.2%) were Buddhist, 15 (10.6%) were Christian, and 9 (2.1%) were others.

Reliability analysis

Table 1 showed all the items had corrected-item total correlation more than 0.3. Thus all the items were maintained in the questionnaire. The Cronbach's alpha value for the Malay version GHQ-12 was 0.852. This analysis suggested that the items of the GHQ-12 were reliable as having high internal consistency.

Sensitivity and specificity analysis

Table 2 and Figure 1 showed that GHQ score at 3/4 have the optimum sensitivity and specificity which were 81.3% and 75.3% respectively with positive predictive value of 62.9% as well as having area under curve more than 0.7. The analysis showed that the GHQ-12 was considered as having an acceptable predictive and discriminative value in detecting distressed medical students.

Table 1: Reliability analysis on the questions of English-Malay version GHQ-12.

No.	Statements	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q1	*Kekurangan tidur kerana risau?* (Lost much sleep over worry)	0.400	0.850
Q2	*Sentiasa merasa tertekan/tegang?* (Felt constantly under strain?)	0.519	0.841
Q3	*Boleh menumpukan perhatian kepada apa sahaja yang dibuat?* (Been able to concentrate on what you are doing?)	0.539	0.840
Q4	*Rasa yang anda memainkan peranan yang berguna dalam banyak perkara?* (Felt that you are playing useful part in things?)	0.331	0.853
Q5	*Dapat mengatasi masaalah-masaalah anda?* (Been able to face up to your problem?)	0.573	0.838
Q6	*Merasa mampu membuat keputusan tentang sesuatu?* (Felt capable of making decisions about things?)	0.537	0.840
Q7	*Rasa yang tidak dapat mengatasi kesukaran/ masalah anda?* (Felt you could not overcome your difficulties?)	0.514	0.841
Q8	*Rasa cukup gembira dalam segala hal yang*	0.639	0.832

	difikirkan? (Been feeling reasonably happy, all things considered?)		
Q9	*Dapat menikmati kegiatan harian anda?* (Been able to enjoy your normal day to day activities?)	0.541	0.839
Q10	*Merasa tidak gembira dan sedih?* (Been feeling unhappy or depressed?)	0.629	0.833
Q11	*Telah hilang kepercayaan pada diri anda sendiri?* (Been losing confidence in yourself?)	0.519	0.841
Q12	*Memikirkan diri anda seorang yang tidak berguna?* (Been thinking of yourself as a worthless person?)	0.567	0.839
Total Alpha = 0.852			

Table 2: The area under ROC curve, sensitivity, specificity, positive and negative predictive values of different cut-off points for significant distress (detection based on BDI-II).

GHQ score	The Area Under ROC curve	Sensitivity %	Specificity %	PPV %	NPV %
0/1	0.66	95.8	36.6	43.8	94.4
1/2	0.72	89.6	54.8	50.6	91.1
2/3	0.77	87.5	66.7	57.5	91.2
3/4	0.78	81.3	75.3	62.9	88.6
4/5	0.72	60.4	83.9	65.9	80.4
5/6	0.72	56.3	87.1	69.2	79.4
6/7	0.64	39.6	89.2	65.5	74.1
7/8	0.61	27.1	94.6	72.2	71.5
8/9	0.59	20.8	96.8	76.9	70.3
9/10	0.58	16.7	98.9	88.9	69.7
10/11	0.55	10.4	98.9	83.3	68.1
11/12	0.54	2.3	100.0	100.0	66.4

ROC = Receiver Operating Characteristics, **PPV** = Positive Predictive Value,
NPV = Negative Predictive Value

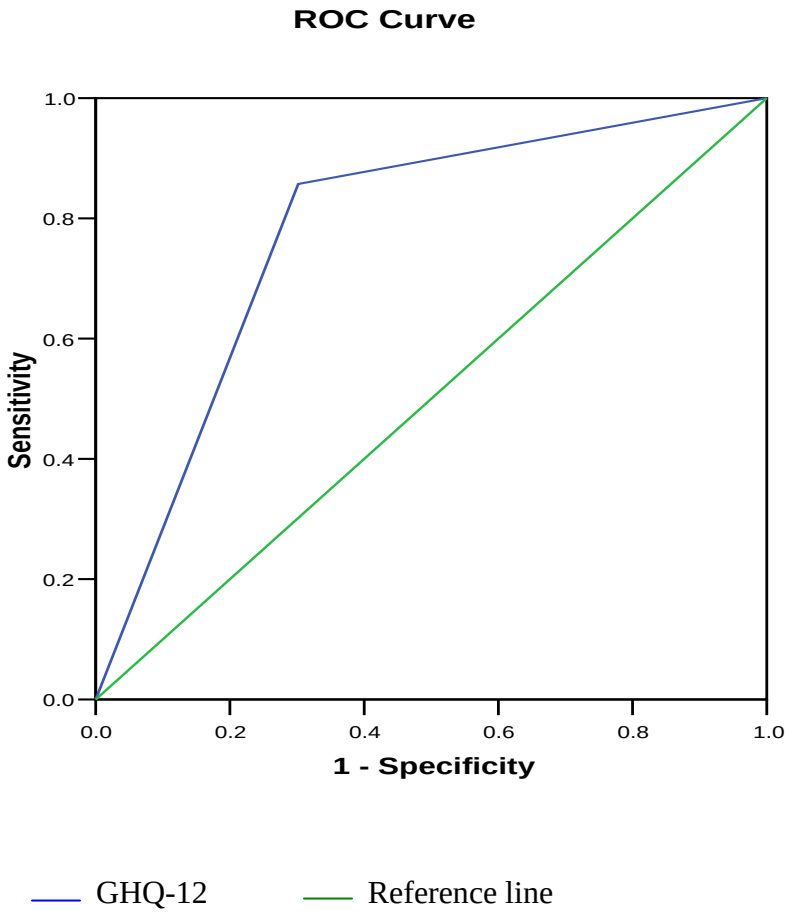

Figure1: The area under ROC curve of the GHQ score at cut-off points 3/4 for significant distress (diagnoses were based on the BDI-II).

Discussion

Reliability generally is defined as consistency or reproducibility of measurement over time or occasions, whereas validity generally is defined as to what extent the measurement measures what it should measure [12, 15, 16]. Sensitivity is defined as the proportion of persons with disease who test positive, whereas, specificity is defined as the proportion of persons without disease who test negative [14-16]. Therefore sensitivity and specificity describe how well the test predicts and discriminates between patients with and without disease. The accuracy of a test depends on how well the test separates the group being tested into those with and without the disease in question. Accuracy is measured by the area under the ROC curve [14]. The purpose of present study is to determine the sensitivity, specificity and internal consistency of the Malay version GHQ-12 which could be used in detecting distressed medical students. Cronbach's alpha value is commonly used by researchers in determining the internal consistency of an instrument, whereas ROC

curve analysis is used to determine the specificity and sensitivity. In this study, the same analysis was applied to determine the sensitivity, specificity and internal consistency of the Malay version GHQ-12.

ROC curve analysis has shown that the optimum cut-off point to detect distress was 3/4 as shown in Table 1. The analysis has also shown the GHQ score of 3/4 have acceptable predictive and discriminative values as the sensitivity, specificity and area under ROC curve was more than 0.7. It reflected the ability of GHQ-12 to discriminate between distressed and non-distressed medical students. The findings were evidence to support and suggest that the GHQ-12 was a valid instrument to detect distressed medical students. It was noteworthy that present study using the BDI-II as a reference group to compare with the tested instrument which was not the gold-standard method, which usually done through clinical-structured-interview by psychiatrists, in determining the specificity and sensitivity of the instrument. However, present study finding is comparable with the Goldberg et al (1997) finding which yielded sensitivity and specificity about 83.7% and 79.0% respectively.

Findings from reliability analysis suggested that all the 12 items have corrected-item total correlation value more than 0.3 as shown in Table 2; Therefore all the items were maintained in the GHQ-12. All the items have shown a measure of high internal consistency as having Cronbach's alpha value more than 0.7 as shown in Table 2; it reflected the reliability of the GHQ-12. The findings were evidence to support and suggest that the GHQ-12 is a reliable instrument that could be used in the future to detect distressed medical students.

Conclusion

This study showed the Malay version GHQ-12 is a valid and reliable screening tool in detecting distressed medical students. The optimum cut-off point of the GHQ-12 to detect distressed medical students was 3/4.

Reference

1. Abdul Hamid A.R, Hatta S.M.A (1996), Validation study of the Malay language translation of General Health Questionnaire (GHQ-30), Malaysian Psychiatry, 2 (4), 34-37.

2. Maniam T. (1996). Validation of the General Health Questionnaire (GHQ-30) for a Malaysian population, Mal J Psych, 4 (2): 25-31.

3. Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O. & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med, 27(1), 191-197

4. Jackson C (2007). The General Health Questionnaire, Occupational Medicine, 57, 59.

5. Montazeri A., Harrichi A.M., Shariati M., Garmaroudi G., Ebadi M., & Fateh A. (2003). The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version, Health and Quality of Life Outcomes, 1 (66).

6. Quek KF, Low WY, Razack AH, Loh CS (2001). Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian Study, Psychiatry and Clinical Neurosciences, 55, 509-513.

7. Nagyova I., Krol B., Szilasiova A., E. Stewart R., P. van Djik J., van den Heuval W.J.A. (2000). General Health Questionnaire-28: psychometric evaluation of the Slovak version , Studio Pysholoiga, 42 (4), 351-361.

8. Banks M H (1983) Validity of the General Health Questionnaire in a young community sample, Pyschological Medicine, 13, 349-54.

9. Radanovic Z, Eric L J (1983) Validity of the General Health Questionnaire in Yugoslav student population, Psychological Medicine, 13, 205-7.

10. Goldberg D (1978). Manual of the General Health Questionnaire. NFER Publishing Company.

11. Costello, Anna B & Jason Osborne (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, Practical Assessment Research & Evaluation, 10 (7). http://pareonline.net/getvn?v=10&n=7. Accessed on August 2008.

12. Downing S.M. Reliability: on the reproducibility of assessment data, Medical Education, 2004; 38: 1006-1012.

13. Wan Mohd Rushidi Wan Mahmud, Amir Awang, Iran Herman, Mahmood Nazaar Mohamed (2004). Analysis of the psychometric properties of the Malay version of Beck Depression Inventory-II (BDI-II) among postpartum women in Kedah, north west of peninsular Malaysia, Medical Journal of Medical Sciences, 11(2), 19-25.

14. Thomas G.T. Interpreting diagnostic test, University of Nebraka Medical Center. Available online: http://gim.unmc.edu/dxtests/Default.htm. Accessed on March 2009.

15. Indrayan A. & Satyanarayana L. Biostatistics for medical, nursing and pharmacy students, Prentice-Hall of India, New Delhi, 2006.

16. Thomas K. Crowl. Fundamentals of educational research (2nd ed.), McGraw-Hill, United States of America, 1996.

Corresponding Author*: Muhamad Saiful Bahri Yusoff, Lecturer, Medical Education Department School of Medical Sciences, 16150 Universiti Sains Malaysia.*

Email: msaiful@kb.usm.my

Received: 21 September 2009 Accepted: 13 October 2009