ORIGINAL ARTICLE

Reliability and Validity of the Malay Version of the General Health Questionnaire (GHQ-12) Among Urological Patients

K F Quek, M.Phil*, W Y Low, PhD, AFBPsS*, A H Razack, FRCS, C B Chua, FRCS (Urol)**, C S Loh, FRCS (Urol)*****

*Health Research Development Unit, **Department of Surgery, University Malaya Medical Centre, 50603 Kuala Lumpur, ***Consultant Urologist, Sunway Medical Centre & Gleneagles Intan Medical Centre, Selangor

Summary

The aim of the study was to validate the Malay version of the General Health Questionnaire (GHQ-12) in patients with psychiatric morbidity secondary to urological disorder. Validity and reliability were studied in patients with lower urinary tract symptoms (LUTS) and patients without LUTS. Internal consistency was excellent. A high degree of internal consistency was observed for each of the 12 items and total scores (Cronbach's alpha value = 0.50 and higher and 0.65 respectively. Test-retest correlation coefficient for the 12 items scores was highly significant. Intraclass correlation coefficient was high (ICC=0.47 and above). A significant level between baseline and post-treatment scores were observed across 3 items in the surgical group. The Mal-GHQ-12 is a suitable, reliable, valid and sensitive to clinical change in the Malaysian population.

Key Words: General Health Questionnaire (GHQ-12), Cronbach's alpha, intraclass correlation coefficient, internal consistency, test-retest reliability and validity

Introduction

The 12-item General Health Questionnaire (GHQ-12) has become a commonly used instrument in multicentre, international clinical trials designed and widely used to detect non-psychotic psychiatric disorders[1]. The GHQ-12 developed by Goldberg is a multidimensional tool and has been widely used in the studies of psychiatric morbidity in many countries both in the community-based study as well as the clinical studies. The GHQ-12 is an instrument in identifying states of psychiatric morbidity, anxiety and depression. It detects on the form of psychiatric disorder which may be relevant to a patient's presence in a medical clinic and focuses on psychological compounds of ill health[2]. Factors such as the presence of physical illness, comorbidity with other psychiatric disorders and the presence of cognitive impairment which are more frequent in the elderly can influence its discriminating ability[3].

This article was accepted: 3 March 2002
Corresponding Author: Low Wah Yun, Health Research Development Unit, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

The presence of somatic symptoms, irrespective of aetiology, was associated with increased social and psychiatric morbidity across disparate culture [4]. The validation of GHQ-12 has been established widely in different countries [5].

GHQ-12 has been used in detecting psychiatric disorder secondary to urological disorder such as benign prostatic hyperplasia (BPH) [6]. Lower urinary tract symptoms (LUTS) from BPH is a common disorder which may affect the patient's perceived quality of life (QOL) in psychological aspects such as psychiatric morbidity, anxiety, and depression [6]. LUTS can be very bothersome for patients and interference with daily activities, and especially, symptoms such as frequency, nocturia, urgency, urge incontinence and dribbling can psychologically disturb the patients' quality of life [6-8].

Existing self-report measures of psychiatric disorders such as the General Health Questionnaire (GHQ-12) has been widely used and developed in Western countries and therefore a translated version of GHQ-12 is needed for the local population for psychiatric disorder assessment. The 12-item GHQ-12 is easy to administer, simple, short and only requires less than 7 minutes to be filled by respondents [1].

The present study was conducted at University Malaya Medical Centre Kuala Lumpur and was designed to assess the reliability and validity of the Malay version of General Health Questionnaire (Mal-GHQ-12) among urological patients.

Materials and Methods

The Mal-GHQ-12 was translated using the back translation technique[9]. The patients were selected based on the inclusion and exclusion criteria set in this study. Patients with LUTS were chosen because it was found that LUTS contributed to psychological effects among the patients. For patients with LUTS in the surgical group, the inclusion criteria were patients must in stable clinical condition and literate (understand and able to answer), and able to give informed consent. Patients who were unable to read, write or understand were excluded in the study, as were patients with any major medical history and who were physically disabled. For the control group of patients, the inclusion criteria includes patients who were free from all major chronic and acute diseases and those with renal stones with minimal severity who did not seek treatment for LUTS, while the exclusion criteria were those seeking treatment for urological problems which include BPH and urinary tract infections (UTI). Renal stones patients were taken as control group in order to have a closer urologic condition for comparison. All patients in the surgical group (TURP) were patients who had TURP from various indications which include acute and chronic retention, recurrent UTI, severe symptoms, failure of medical treatment and renal impairment secondary to bladder outlet obstruction. The study protocol was approved by the Ethics Committee, University Malaya Medical Centre, Kuala Lumpur.

Study sample

The psychometric properties of the Mal-GHQ-12 were assessed in two different groups, namely patients with LUTS undergoing surgical treatment (TURP)(n=20), and a control group of patients (renal stones)(n=20). Validity and reliability were studied in a control group of patients with renal stones and sensitivity of change was assessed in a group of patients who were diagnosed of LUTS admitted for transurethral resection of the prostate (TURP). Management decisions was entirely done by a Urologist (A.H.R.) in all cases based on clinical criteria such as medical history, physical and rectal examinations.

Data Collection

All patients in the surgical group (TURP) and control group of patients gave their written informed consent to participate in the study after explanation was given on the purpose and nature

of the study. Patients then completed the Mal-GHQ-12. All questionnaires were self-administered although assisted guidance was available by one of the authors (K.F.Q) of this study to assist the patients where necessary. All patients included in the validity study were re-tested at 12 weeks after the first administration of the Mal-GHQ-12 (baseline). In the sensitivity to change study, patients completed the questionnaires one-week before the surgical treatment and were re-tested at three months after TURP.

Data Analysis

Cronbach's alpha coefficient[10] was used to assess the internal consistency of the Mal-GHQ-12. The internal consistency shows the resulting values of Cronbach's alpha for the scale when individual items are excluded from the analysis. Intraclass correlation coefficient (ICC) which is derived from analysis of variance was used to assess the test-retest reliability. The values of ICC vary from 1 (perfectly reliable) to 0 (totally unreliable)[11]. Sensitivity to change of the Mal-GHQ-12 was assessed in patients undergoing TURP. Sensitivity to change was analyzed by calculating the difference between Mal-GHQ-12 before and after TURP and dividing it by the standard deviation of the scores between Mal-GHQ-12 before and after TURP (effect size) [12].

The mean differences of Mal-GHQ-12 scores before and after TURP was tested for each individual item and total scores by means of a paired t test or by dividing it by the standard deviation of stable LUTS patients scores (control group) (Guyatt statistic) [13].

Sensitivity and specificity of Mal-GHQ-12 was evaluated by comparing the change between baseline and end point scores following treatment. Sensitivity is the proportion of cases correctly identified whereas specificity is the proportion of non-cases correctly identified. In this study, sensitivity of the Mal-GHQ-12 was assessed by comparing between the mean of pre-treatment and post-treatment of patients who had undergone TURP whereas specificity was assessed by comparing the baseline and 3 months in the control group (renal stones).

Results

A total of 40 respondents were involved in the validity and reliability study. The mean age of the patients in the surgical group was 70.65 (SD=7.31 years) and control group was 49.90 (SD=12.03 years). Patients with LUTS showed higher Mal-GHQ-12 scores.

Internal consistency for the Mal-GHQ-12 was high for all the items of GHQ-12 indicating a high level of homogeneity among items in the scale. Test-retest reliability was assessed in 20 patients after 12 weeks interval and the total Mal-GHQ-12 had an ICC of 0.67 ($p<0.001$) (Table I).

Table II shows pre and post scores, mean difference, effect size, and the Guyatt statistic for individual items, and the total scores. The mean pre intervention score on the total Mal-GHQ-12 was 12.50 (SD=4.49) while the mean post intervention score was 9.70 (SD=2.47) ($p<0.001$), giving an average improvement after TURP on the psychiatric morbidity of 2.80. Overall effect size and therefore sensitivity to change were found to be high indicating TURP-induced urinary symptoms improvement in these patients.

Majority of items of the Mal-GHQ-12 demonstrated a high degree of sensitivity and specificity to the effects of treatment (Table III). Magnitude changes were observed across in half of the items in the TURP group. The lowest magnitude of change was noted in item 3 and 4. In contrast, small magnitude of change was observed in the treatment of the control.

Discriminant validity or ability of the Mal-GHQ-12 scale to discriminate reliably between the surgical treated group and control group was assessed. However, as shown in Table IV, significant

Table I: Validity and Reliability: Mean Test-Retest score, interclass correlation coefficient and internal consistency for individual GHQ-12 items (control)

GHQ-12 Items	Intraclass Correlation Coefficient* (ICC)	Internal Consistency**	Mean Test Score	SD	Mean Retest Score	SD	Mean Difference***	SD
1	0.87	0.86	1.30	0.47	1.30	0.47	0.00	0.32
2	0.83	0.83	0.95	0.76	0.85	0.67	0.01	0.55
3	0.67	0.70	1.25	0.44	1.10	0.31	0.15	0.37
4	0.64	0.62	1.10	0.31	1.10	0.31	0.00	0.32
5	0.84	0.83	0.80	0.70	0.75	0.64	0.05	0.51
6	0.83	0.83	0.70	0.66	0.65	0.67	0.05	0.51
7	0.47	0.47	1.05	0.39	1.15	0.37	0.10	0.45
8	0.50	0.50	1.10	0.31	1.15	0.37	0.05	0.39
9	0.84	0.83	0.75	0.79	0.70	0.80	0.05	0.60
10	0.91	0.91	0.50	0.69	0.55	0.69	0.05	0.39
11	0.92	0.92	0.40	0.60	0.40	0.60	0.00	0.32
12	0.79	0.79	1.10	0.31	1.05	0.22	0.05	0.22
Total Scores	0.67	0.65	10.95	5.07	9.25	3.21	1.70	5.16

P<0.001 for all ICCs(*)

Cronbach's alpha: note that Cronbach's alpha value given for each item represent the effect of removing that item from the calculation of the alpha value eg. if item 1 is ommitted, the resulting value for the scale is 0.86, if item 2 is omitted, it is 0.83, and so forth) (**)

t test for paired comparisons not significant (***)

Table II: Sensitivity to change: mean scores before and after TURP, effect size and Guyatt Statistic

GHQ-12 Items	Pre TURP		Post TURP					
	Mean	SD	Mean	SD	Mean Differences	SD	Effect size	Guyatt statistic
1	1.35	0.59	1.10	0.45	0.25	0.64	0.42	0.53
2	1.15	0.67	0.80	0.52	0.35	0.81	0.52	0.46
3	1.20	0.41	1.05	0.22	0.15	0.37	0.37	0.34
4	1.20	0.41	1.05	0.22	0.15	0.37	0.37	0.48
5	0.75	0.55	0.80	0.70	0.05	0.76	0.09	0.07
6	1.30	0.73	0.70	0.57	0.60*	0.75	0.82	0.91
7	1.40	0.50	1.05	0.51	0.35*	0.59	0.70	0.90
8	1.25	0.44	1.10	0.45	0.15	0.49	0.34	0.48
9	1.05	0.61	0.70	0.86	0.35	0.99	0.57	0.44
10	0.55	0.51	0.45	0.69	0.10	0.85	0.20	0.14
11	0.30	0.47	0.20	0.52	0.10	0.79	0.21	0.17
12	1.15	0.49	1.05	0.22	0.10	0.55	0.20	0.32
Total Scores	12.50	4.49	9.70	2.47	2.80*	3.50	0.62	0.55

Effect size = Mean different / SD PreTURP
Guyatt statistic = Mean different / SD of stable LUTS patient (control group)
* t test for paired comparisions significant

Table III
GHQ-12 items characteristics of patient undergoing TURP and the control group
Sensitivity and Specificity

Sensitivity	n	Mean Changes	SEM	t statistic	p value
Items					
1	20	0.25	0.14	1.75	0.1
2	20	0.35	0.18	1.93	0.1
3	20	0.15	0.08	1.83	0.08
4	20	0.15	0.08	1.83	0.08
5	20	0.05	0.17	0.29	0.77
6	20	0.60	0.17	3.56	0.002
7	20	0.35	0.13	2.67	0.15
8	20	0.15	0.11	1.37	0.186
9	20	0.35	0.22	1.58	0.13
10	20	0.10	0.19	0.52	0.61
11	20	0.10	0.18	0.57	0.58
12	20	0.10	0.12	0.81	0.43
Total scores	20	2.80	0.78	0.57	0.005

Sensitivity	n	Mean Changes	SEM	t statistic	p value
Items					
1	20	0.00	0.01	0.00	1
2	20	0.01	0.12	0.81	0.43
3	20	0.15	0.01	1.83	0.08
4	20	0.00	0.01	0.00	1
5	20	0.05	0.11	0.44	0.67
6	20	0.05	0.11	0.44	0.67
7	20	0.10	0.10	1.00	0.33
8	20	0.05	0.01	0.57	0.58
9	20	0.05	0.13	0.37	0.72
10	20	0.05	0.01	0.57	0.58
11	20	0.00	0.01	0.00	1
12	20	0.05	0.01	1.00	0.33
Total scores	20	1.70	1.15	1.47	0.16

Table IV
GHQ-12 items characteristics: Discriminant validity

Items	Pre TURP patient		Control (Baseline)						
	Mean	SEM	Mean	SEM	Mean Differences	SEM	95% Confidence Lower	Internal Higher	p value
1	1.35	0.13	1.30	0.10	0.05	0.17	0.29	0.39	0.77
2	1.15	0.15	0.95	0.17	0.20	0.23	0.26	0.66	0.38
3	1.20	0.09	1.25	0.10	0.05	0.13	0.32	0.22	0.71
4	1.20	0.09	1.10	0.07	0.10	0.11	0.13	0.33	0.39
5	0.75	0.12	0.80	0.16	0.05	0.20	0.45	0.35	0.80
6	1.30	0.16	0.70	0.15	0.60	0.22	0.15	1.04	0.01
7	1.40	0.11	1.05	0.09	0.35	0.14	0.05	0.64	0.02
8	1.25	0.10	1.10	0.07	0.15	0.12	0.06	0.40	0.22
9	1.05	0.13	0.75	0.18	0.30	0.22	0.10	0.75	0.18
10	0.55	0.11	0.50	0.15	0.05	0.19	0.15	0.44	0.80
11	0.30	0.10	0.40	0.13	0.10	0.17	0.44	0.24	0.56
12	1.15	0.11	1.10	0.07	0.05	0.13	0.21	0.31	0.70
Total Scores	12.50	1.00	10.95	1.13	1.55	1.51	1.52	4.62	0.31

differences were observed for only two items between the surgical group and the control group.

Discussion

The Mal-GHQ-12 has proven to be a valid and a reliable tool in assessing psychiatric morbidity patients with LUTS although the findings were not so obvious. Its ability to discriminate between patients with LUTS and those without showed high levels of sensitivity and specificity. The effect size obtained indicated a high degree of sensitivity to change. Similar findings were also noted in other studies [12,13].

These findings provide substantial assurance that scores obtained using the Mal-GHQ-12 are reliable. The present findings on the reliability and validity were consistent with scores of GHQ-12 validated in various countries [14].

The psychometric properties of the Mal-GHQ-12 validated in this study and GHQ-12 validated in other countries showed that they are virtually identical with respect to their measurement properties[13]. No statistical significance was noted in the test-retest reliability.

Conclusion

The reliability and validity of the Mal-GHQ-12 for 40 urological patients were tested. The intraclass correlation coefficient for each items and total scores of GHQ-12 indicated high intraclass reliability. Likewise the high Cronbach's alpha coefficient indicates that Mal-GHQ-12 exhibited a good internal consistency. Thus Mal-GHQ-12 is a useful and accurate tool for assessing the severity of psychiatric morbidity secondary to LUTS among urological patients.

Acknowledgements

The authors would like to express their gratitude to the University Malaya's Research and Development Management Unit for providing the research grant for this study. The authors also thank Dewan Bahasa and Pustaka, Kuala Lumpur for the verification of the questionnaire.

References

1. Goldberg D: General Health Questionnaire (GHQ-12). The NFER-NELSON Publishing Company Ltd, Darville House, 2 Oxford Road East, Windsor, Berkshire, 1978.
2. Goldberg DP, Williams P: User's Guide to the General Health Questionnaire. Windsor. NFER-Nelson. 1988.
3. Papassotiropoulos A, Heun R, Maier W: Age and cognitive component influence the performance of the general health questionnaire. Compr Psyc 1997; 38(6): 335-40.
4. Kisely S, Goldberg D, Simon G: A comparison between somatic symptoms with and without clear organic cause: results of an international study. Psychol Med 1997; 27(5): 1011-19.
5. Goldberg D: Identifying psychiatric illness among general medical patients. BMJ 1985; 291: 161-62.
6. Quek KF, Low WY, Razack AH, Loh CS: Psychological effects of treatment for lower urinary tract symptoms. BJU International 2000; 86: 630-33.
7. Lukacs B, Leplege A, Mc Carthy C, Comet D: Construction and validation of a BPH specific health related quality of life scale, for medical outcome research studies. In Cockett ATK, Khoury S, Aso Y et al., eds, Proceedings of the Second International Consultation on Benign Prostatic Hyperplasia (BPH). Paris. SCI. 1994. pp 139-43.
8. Barry MJ, Fowler FJ Jr, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett ATK: The American Urological Association symptom index for benign prostatic hyperplasia. J Urol 1992; 148: 1549-57.
9. Brislin RW: Back-translation for cross-cultural research. J Cross-Cultural Psychol 1970; 1: 185-216.
10. Cronbach LJ: Coefficient alpha and the internal structure of test. Psychometrika 1951; 16: 297.
11. Deyo RA, Dichr P, Patrick DL: Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clinical Trials 1991; 12 (supp 1): 142S-158S.
12. Cohen J. Statistical Power analysis for the Behavioral Sciences. Academic Press, New York, 1977. pp75-105.
13. Guyatt G, Walter S, Norman G: Measuring changes overtime: assessing the usefulness of evaluative instruments. J Chronic Dis 1987; 40: 171-78.
14. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinell, OG, Rutter C: The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997; 27: 191-97.