Scientific Research and Essays Vol. 6(2), pp. 278-282, 18 January, 2011
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE10.638
ISSN 1992-2248 ©2011 Academic Journals

Full Length Research Paper

Reliability and validity of patient health questionnaire: Depressive syndrome module for outpatients

Cuidong Bian[1], Chunbo Li[2], Qianglin Duan[3]* and Heng Wu[4]

[1]Department of Urology, Tongji Hospital of Tongji University, Shanghai 200065, China.
[2]Shanghai Mental Health Center, Shanghai 200030, China.
[3]Department of Scientific Research, Tongji Hospital of Tongji University, Shanghai 200065, China.
[4]Department of Psychiatry, Tongji Hospital of Tongji University, Shanghai 200065, China.

Accepted 25 October, 2010

We evaluated the reliability, validity and detection rate of the Depressive Syndrome module of the Patient Health Questionnaire (PHQ-9) in general hospital outpatients. Totally 600 general hospital outpatients were evaluated using the PHQ-9. The internal reliability, test-retest reliability and validity were examined. Cronbach's α coefficient of PHQ-9 was 0.857 and the test-retest reliability was 0.947. The correlation coefficient of the nine items with the total score of the scale was 0.588 - 0.784. The sensitivity, specificity of PHQ-9 and Kappa value was 91, 97% and 0.884, respectively. The detection rate was 16.3% (95% CI: 13.4 - 19.3%). The Chinese version PHQ-9 was shown to have good reliability and validity for screening of depressive syndrome in general hospital outpatients.

Key words: Patient health questionnaire, depressive syndrome module-9, reliability, validity, detection rate.

INTRODUCTION

Depressive disorder is a common psychiatric disease, and approximately 10 - 19% patients visiting primary health care organizations suffer from depressive disorders (Spitzer et al., 1999). However, as patients often complain of somatic symptoms rather than mental symptoms, there is a low recognition rate of physicians for depressive disorder. Less than 60% of cases can be effectively cured (Henriques et al., 2009). Patients that are not effectively treated often develop serious functional impairment and long-term chronic somatic damage, and their suicide rates are also increased, leading to excessive consumption of medical resources (McQuaid et al., 1995; Ormel et al., 1998; Schonfeld et al., 1997). Selection of correct and effective screening instruments is important to improve the detection rate of mental disorders (Löwe et al., 2004). Currently, some commonly-used international screening questionnaires are designed on different focuses, so they have certain restrictions in clinical application. Spitzer et al. (1999) developed a self-administered questionnaire for primary care - the Patient Health Questionnaire (PHQ) according to DSM-IV diagnostic criteria. In the questionnaire, the depression scale PHQ-9 is developed based on the 9 depressive symptoms in DSM-IV. Because of its briefness as well as easy operation and scoring, the PHQ-9 is quickly applied in a wide range of scientific research and clinical practice. This study aimed to test the reliability, validity and detection rate of the Chinese version of PHQ-9, and to investigate its application value in outpatients from general hospitals.

METHODS

Subjects

A questionnaire survey on outpatients in the departments of internal medicine, gynecology, mammary gland and general surgery of Tongji Hospital was conducted from October, 2007 to March, 2008. The inclusion criteria included age ≥ 18 years, primary education or higher, and patients without serious physical diseases, verbal and hearing disorders. Informed consent was obtained from all the subjects.

*Corresponding author. E-mail: duan09490@hotmail.com.

Table 1. Correlation of inter-items of PHQ-9 and each item with the total score.

Items	1	2	3	4	5	6	7	8	9
1	1								
2	0.718**	1							
3	0.480**	0.475**	1						
4	0.533**	0.505**	0.412**	1					
5	0.325**	0.312**	0.236**	0.401**	1				
6	0.407**	0.457**	0.351**	0.449**	0.257**	1			
7	0.548**	0.544**	0.461**	0.520**	0.269**	0.297**	1		
8	0.302**	0.342**	0.333**	0.448**	0.505**	0.290**	0.312**	1	
9	0.617**	0.625**	0.479**	0.420**	0.418**	0.331**	0.443**	0.397**	1
Total score	0.783**	0.784**	0.685**	0.761**	0.588**	0.626**	0.702**	0.627**	0.711**

Note: ** P <0.01 (two-tailed test).

Tools

With the consent of the original authors of PHQ (Spitzer R, Kroenke K), the project director (Chunbo Li) introduced the questionnaire, and after repeated translation and back translation by three professional researchers and a English-proficient psychiatric chief physician, a final version for research was determined. PHQ-9 has nine items, measuring how long patients suffered from nine problems including interest deficiency, depressed mood in the past two weeks. Answer of "not at all" "several days" "More than half the days" and "almost every day" scores 0, 1, 2 and 3, respectively. The PHQ-9 score ranges from 0 to 27. Score 5, 10, 15, and 20 represent the thresholds for mild, moderate, moderately severe, and severe depression, respectively.

The Hospital Anxiety and Depression Scale (HADS) and the Hamilton Depression Rating Scale (Hamilton Depression Scale, HAMD) were used as tools for validity test. The depression screening part of Structured Clinical Interview for DSM-IV (SCID) was used as the "gold standard" for diagnoses.

Assessment methods

Cluster random sampling was made with departments as sampling units. Using the defined instructions, researchers guided the selected outpatients to complete the self-administered scales PHQ-9 and HADS, and made assessment on HAMD. A total of 776 patients were evaluated, of which 172 refused to response and 4 quit half-way, and 600 valid questionnaires were obtained. Those who refused to response and quit predominantly were surgical outpatients, due to short waiting times, patients having no time to complete the questionnaire or patients refused to fill their general personal information. Forty-four patients selected from the 600 patients were phone-administered for PHQ-9 test-retest 7 - 14 days after the completion of the initial assessment. Thirty-two patients selected from the samples with PHQ-9 ≥10 and 64 patients from PHQ-9 <10 were reviewed by two psychiatric physicians for SCID, and the reliability and validity study on PHQ-9 was performed.

Statistical analyses

All data were input into a database established by EPIDATA 3.1, and statistically analyzed by SPSS 13.0 statistical software package. Based on data distribution characteristics, statistical methods including descriptive statistics, chi-square test, reliability analysis, Pearson product-moment correlation analysis were used.

RESULTS AND DISCUSSION

General data

Among the 600 patients that completed the PHQ-9, there were 273 (45.5%) males and 327 (54.5%) females with a mean age of 51.5 ± 16.4 years (range: 18 - 77 years). Education component included 109 (18.2%) of primary school, 196 (32.7%) junior high, 160 (26.7%) senior high school, 135 (22.4%) universities or higher. Marital status showed that 503 (83.8%) were married, 42 (7.0%) were widowed or divorced, and 55 (9.2%) were single. A total of 568 (94.7%) responders were PHQ-9 <10, and 32 (5.3%) were PHQ-9 ≥ 10.

Reliability study

The cronbach's α coefficient was 0.857. The correlation coefficients of each item with the total scores of the scales were 0.588 - 0.784, and the correlation coefficients of inter-items were 0.236 - 0.718, all of which had remarkable statistical significance (P <0.01) (Table 1).

Test-retest reliability was elevated. The correlation between the two tests was calculated by Pearson product-moment correlation analysis, and the correlation coefficient between the total scores of the two measurements was 0.947, with remarkable statistical significance (P <0. 01).

Validity study

Correlation analysis among PHQ-9 total scores, HADS total scores, HADS anxiety subscale scores, HADS depression subscale scores and HAMD total scores was conducted, in order to examine the PHQ-9 criterion validity (Table 2). MIn the 32 patients with PHQ-9 ≥ 10, 30 were in line with the SCID diagnostic criteria for depressive disorder and in the 64 patients with PHQ-9

Table 2. Correlation analysis between PHQ-9 with HADS and HAMD.

	HADS total score	HADS anxiety score	HADS depression score	HAMD total score
PHQ-9 total score	0.789**	0.550**	0.792**	0.811**

Note: ** P <0.01 (two-tailed test).

Table 3. Validity comparison between PHQ-9 and SCID.

	PHQ-9(+)	PHQ-9(-)	Total
SCID(+)	30	3	33
SCID (-)	2	61	63
Total	32	64	96

<10, 3 were in line with the SCID diagnostic criteria for depressive disorder. The sensitivity and specificity of PHQ-9 and Kappa value was 91, 97% and 0.884, respectively (P < 0.01) (Table 3).

Detection rate of depressive syndrome

Depressive syndrome was detected in 98 subjects, accounting for 16.3% (95%CI: 13.4 - 19.3%) of the total sample size. Among them, 22 were mild, accounting for 22.4% (95%CI: 14.2 - 30.7%) of the total detected number; 59 were moderate, accounting for 60.2% (95%CI: 50.5 - 69.9%); 17 were severe, accounting for 17.3% (95%CI: 9.9 - 24.8%). Table 4 shows the demographic distribution of the subjects with depressive syndrome detected. Among them, the difference in the gender distribution was statistically significant, and more females than males. The age and employment situation distribution was statistically significant. Differences in education level and marital status distribution did not reach statistical significance.

Discussion on reliability

PHQ-9 is the most commonly used scale in PHQ in research setting, and has shown a good psychometric property in various studies. For instance, the use of the PHQ-9 on the university students in Nigeria showed that the internal consistency (Cronbach's α coefficient) of each item in the PHQ-9 was 0.85, and had a good test-retest reliability over one month interval (r = 0.894, P < 0.001) (Abiodun et al., 2006).

This study showed that Cronbach's α coefficient of the PHQ-9 was 0.857, suggesting that the PHQ has good internal consistency. The correlation coefficient of the total score of PHQ-9 measured twice within two weeks was 0.947, suggesting that the PHQ has good cross-time stability. Analysis on the items of the scale showed that the correlation coefficients of each item with the total score of the scale were 0.588 - 0.784, and the correlation coefficients of inter-items were 0.236 - 0.718, all of which had remarkable statistical significance (P <0.01), indicating that PHQ-9 has a good internal homogeneity. Both Löwe et al. (2006) demonstrated that the PHQ-9 had good reliability. To sum up, it can be concluded that the PHQ-9 also has a good reliability in general hospital application.

Discussion on validity

Validity mainly refers to the accuracy of measuring instruments on questions to be assessed, including content validity, construct validity and criterion validity, etc. The study by Martin et al. (2006) confirmed that the PHQ-9 depression scale had good construct validity. In Honduras, a study on the feasibility of a Spanish version of PHQ-9 depression scale for depression screening showed that the sensitivity of the PHQ-9 was 77% and the specificity was 100% (Wulsin et al., 2002). The application of PHQ-9 for patients with post-traumatic stress disorder showed that when the PHQ-9 ≥ 10, the sensitivity was 91% and the specificity was 89% (Williams et al., 2005).

This study used the HADS, HAMD and SCID as the tools for testing the criterion validity of the PHQ. Construct validity was supported by significant correlations with measures of similar constructs. Factor analysis confirms multiple dimensions of eating disorder symptoms but suggests possible culture-specific variation in this population. According to views on the construct validity, scales with the same characteristics or structure should have a high inter-scale consistency or correlation, and the test on the consistency and correlation analysis between two types of scales is the congruent validity (Jiang, 1999). This study revealed that the PHQ-9 has a high correlation with the criterion HAMD and the HADS of the same characteristics, therefore, it can be concluded that PHQ has a good concurrent validity and congruent validity with the HADS and HAMD.

The results of this study showed that the sensitivity,

Table 4. Demography of patients with depressive syndrome.

		Patients in line with depressive syndrome (N)	Total sample size (N)	Percentage of the total sample size (%)	95%CI	*P* value of chi-square test
Total		98	600	16.3	13.4 - 19.3	
Gender	Male	29	273	10.6	7.0 - 14.3	0.001
	Female	69	327	21.1	16.7 - 25.5	
Age	<50	31	262	13.4	9.0 - 17.8	0.002
	50-64	41	167	32.5	24.4 - 40.7	
	≥65	26	171	17.9	11.7 - 24.2	
Education level	Primary school	23	109	21.1	13.4 - 28.8	0.083
	Junior high school	38	196	19.4	13.9 - 24.9	
	High school	18	160	11.3	6.4 - 16.1	

specificity of PHQ-9 and Kappa value was 91, 97% and 0.884, respectively ($P < 0.01$), suggesting that the PHQ-9 has a good criterion validity. This is also identical to a large number of studies abroad. In summary, this study concludes that the Chinese version PHQ-9 is shown to have a good validity.

Discussion on detection rate

Studies abroad (Rief et al., 2004) have found that depression and anxiety symptoms have a higher prevalence in females and the elder than in males and the young. Marital status also has a significant impact on the scores of depression scales. Depression scores of the widowed and divorced are increased and the lowest score is in the unmarried. Depression scores also have a strong correlation with the education and income level. The results of this study showed that there was a significant statistical difference in the detection rate of Depressive Syndrome between male and female, suggesting that women are more susceptible to depressive disorder than men.

This study is the first to apply PHQ-9 in general hospitals, and shows that as a screening scale, the PHQ-9 has a good reliability and validity. In addition, studies abroad have shown that the PHQ-9 is superior to the HADS in the diagnosis of depressive disorders (Bagby et al., 2004), which may be because the items of PHQ are in line with the DSM-IV diagnostic criteria and include all of the disorders for diagnosis of depression disorder. Further, as a screening tool, the PHQ-9 in evaluating the severity of depressive disorder also shows good validity (Kroenke et al., 2001). The PHQ-9 is also very sensitive to changes of depression after treatment on patients with depressive disorder (Löwe et al., 2004; Gilbody et al., 2007). Currently, the PHQ-9 is the only self-administered depression scale that are effective in both of screening disease severity and result measurement (Löwe et al., 2004). Therefore, in the United States, the PHQ-9 is recommended as a routine depression scale by primary health care institutions (Lawson et al., 2002; Reynolds, 2010).

The research on the application of PHQ-9 in China is still in the start-up stage. The retrievable relevant literature on PHQ-9 is only the application in the community elderly population (Xu et al., 2005), which also showed good reliability and validity, but its ample size is very small. Therefore, the PHQ-9 has a wide range of application scope in research setting in China, and its generalization needs the mutual efforts of relevant workers.

REFERENCES

Abiodun O, Adewuya, Bola A (2006). Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. J. Affect. Disord., 96:89-93.

Bagby RM, Ryder AG, Schuller DR, Marshall MB (2004). The Hamilton Rating Scale: has the gold standard become a lead weight? Am. J. Psychiatr., 161: 2163-2177.

Gilbody S, Richards D, Brealey S, Hewitt C (2007). Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J. Gen. Intern. Med.22:1596-602.

Henriques SG, Fráguas R, Iosifescu DV, Menezes PR, Lucia MC, Gattaz WF, Martins MA (2009). Recognition of depressive symptoms by physicians. Clinics (Sao Paulo), 64: 629-35.

Kroenke K, West SL, Swindle R, Gilsenan A, Eckert GJ, Dolor R, Stang P, Zhou XH, Hays R, Weinberger M (2001). Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: a randomized trial. J. Am. Med. Assoc., 286: 2947-2955.

Löwe B, Kroenke K, Herzog W, Gräfe K (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J. Affect Disord., 81: 61-66.

Löwe B, Schenkel I, Carney-Doebbeling C, Göbel C (2006). Responsiveness of the PHQ-9 to Psychopharmacological Depression Treatment. Psychosomatics, 47: 62-67.

Martin A, Rief W, Klaiberg A, Braehler E (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen . Hosp. Psychiatr., 28: 71-77.

McQuaid JR, Stein MB, Laffaye C, McCahill ME (1995). Depression in a primary care clinic: The prevalence and impact of an unrecognized

disorder. J. Affect. Disord., 55: 1-10.
Ormel J, Kempen GIJM, Deeg DJH (1998). Functioning, wellbeing, and health perception in late middle-aged and older people:Comparing the effects of depressive symptoms and chronic medical conditions. J. Am. Geriatr. Soc., 46(1) :39-48.
Reynolds WM (2010). The PHQ-9 works well as a screening but not diagnostic instrument for depressive disorder. Evid. Based Ment. Health,13:96.
Rief W, Nanke A, Klaiberg A, Braehler E (2004). Base rates for panic and depression according to the Brief Patient Health Questionnaire:a population-based study, J. Affect Disord., 82: 271-6.
Schonfeld WH, Verboncoeur CJ, Fifer SK, Lipschutz RC, Lubeck DP, Buesching DP (1997). The functioning and well-being of patients with unrecognized anxiety disorders and major depressive disorder. J. Affect Disord., 43: 105-119.
Spitzer RL, Kroenke K, Williams JB (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study J. Am. Med. Assoc. 282: 1737-1744.
Williams LS, Brizendine EJ, Plue L, Bakas T, Tu W, Hendrie H, Kroenke K (2005). Performance of the PHQ-9 as a Screening Tool for Depression after Stroke. Stroke, 36: 635-638.
Wulsin L, Somoza E, Heck J (2002). The Feasibility of Using the Spanish PHQ-9 to Screen for Depression in Primary care in Honduras. Primary Care Companion J. Clin. Psych., 4: 191-195.
Xu Y, Wu HS, Xu YF(2005). The application of patient health questionaire depression scale (PHQ-9) in the community elderly population - validity and reliability analysis. Shanghai Arch. Psychiatr., 19 (5): 257-259.