International Journal of Gerontology xxx (2016) 1–5

Contents lists available at ScienceDirect

International Journal of Gerontology

journal homepage: www.ijge-online.com

Original Article

Validation of the Patient Health Questionnaire for Depression Screening Among the Elderly Patients in Taiwan☆

I-Peng Chen [1], Shen-Ing Liu [1, 2, 3 *], Hui-Chun Huang [3, 4], Fang-Ju Sun [3, 4], Chiu-Rong Huang [3], Ming-Ru Sung [3], Yo-Ping Huang [5]

[1] *Department of Psychiatry, MacKay Memorial Hospital,* [2] *Department of Medicine, MacKay Medical College, New Taipei City, Taiwan,* [3] *Department of Medical Research, MacKay Memorial Hospital,* [4] *MacKay Medicine Nursing and Management College,* [5] *Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan*

ARTICLE INFO

Article history:
Received 4 March 2016
Received in revised form
21 April 2016
Accepted 27 May 2016
Available online xxx

Keywords:
aged,
depression,
primary health care,
reproducibility of results,
surveys and questionnaires

SUMMARY

Background: This study aimed to assess the reliability, validity, and factor structure of the Patient Health Questionnaire-9 (PHQ-9) and its two subscales (PHQ-2, PHQ-1) for screening major depression disorder (MDD) in primary care elderly patients in Taiwan.

Method: A total of 634 participants aged ≥60 years were enrolled from the primary care settings. They completed PHQ-9, and Short Form of Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q-SF). After that, they were interviewed by using Schedule for Clinical Assessments in Neuropsychiatry (SCAN) and 17-item of Hamilton Rating Scale (HAMD-17).

Result: The PHQ-9 demonstrated acceptable internal consistency ($\alpha = 0.77$), test-retest reliability (intraclass correlation coefficient = 0.79), and concurrent validity with HAMD-17 ($r = 0.66$; $p < 0.001$). Exploratory factor analysis yielded 2-factor structure consisted of non-somatic (13.3% variance) and somatic (38.8% variance) dimensions. The correlations between PHQ-9 with Q-LES-Q-SF (whole and two subscales) were moderate (r range −0.47 to −0.53, all $ps < 0.001$). Using SCAN as gold standard, the optimal cutoff score of ≥6 for PHQ-9 demonstrated good sensitivity and specificity for detecting MDD (1.00 and 0.85, respectively). The screening accuracy of the two items version was also acceptable (PHQ-2 score ≥ 2: sensitivity 0.93; specificity 0.90; PHQ-1 score ≥ 1: sensitivity 1.00; specificity 0.77).

Conclusion: The Chinese versions of the PHQ-9 and its two subscales (PHQ-2 and PHQ-1) are valid and reliable to screen MDD in primary care elderly patients in Taiwan and could be proper alternatives to Geriatric Depression Scale 15-item version.

1. Introduction

Depression in old age has become a major public health issue. Despite its high prevalence and association with impairment in quality of life[1], it is frequently under-diagnosed and under-treated, particularly in non-mental health settings[2]. Early recognition and treatment of depression in older population will lead to improve daily functions and mortality[3,4]. Thus, an important strategy to enhance the detection of depression is the use of screening tool.

There are a variety of screening instruments to measure the severity of depression in the different settings, such as the Beck Depression Inventory[5], Centre for Epidemiologic Studies Depression Scale[6], Geriatric Depression Scale 15-item version (GDS-15)[7]. However, these tools have not become very popular in primary care because of the sheer volume of questions listed, which can be quite overwhelming to the patient and time-consuming to complete. A widespread tendency is the use of short versions of scales, simplifying the application of large scales for physicians. The Patient Health Questionnaire-9 (PHQ-9) is half the length of other depression measures, and has been widely in use internationally, including Chinese population[8,9]. The PHQ-9 surveys nine symptoms of the DSM-IV diagnostic criteria for major depressive

☆ Conflicts of interest: All contributing authors declare that they have no conflicts of interest.

* Correspondence to: Professor Shen-Ing Liu, Department of Psychiatry, MacKay Memorial Hospital, No. 45, Minsheng Rd., Tamshui District, New Taipei City 25160, Taiwan, ROC.

E-mail address: maryliuyip@gmail.com (S.-I. Liu).

http://dx.doi.org/10.1016/j.ijge.2016.05.002

disorder (MDD), which means that it can serve as a dual-purpose instrument that evaluate the severity of this disorder and at the same time, establish the MDD diagnosis for the patient[10]. Meanwhile, the ultra-brief screening tools, PHQ-2 (i.e., the first two items of the PHQ-9) and PHQ-1 (single item, depressed mood), have been used as initial screening tests for MDD with the potential to reduce the burden of depression screening.

To our knowledge, PHQ-9 screening characteristics have not been validated with the elderly in primary care in Taiwan. Our aim was to investigate the psychometric properties of the PHQ-9 and its two subscales, the PHQ-1 and PHQ-2, in identifying elderly patients with MDD.

2. Method

2.1. Setting and sample

The original main study was a randomized controlled trial to test the effectiveness of a collaborative care intervention for patients with MDD in primary care in Taipei, which was conducted between January 2009 and March 2012. The pilot trial of the above main study was to test its feasibility and develop appropriate instruments such as PHQ-9[9]. The pilot trial was conducted from September 2007 through December 2008 in six community-based primary care clinics and hospital-based outpatient clinics of the Department of Family Medicine in Taipei.

The institutional review board at the MacKay Memorial Hospital approved all procedures (No. MMH-I-S-241). Eligible participants in the pilot trial were primary care participants who aged ≥18, spoke or understood Mandarin/Fukien dialects, and were absence of severe cognitive or hearing impairment. We chose patients who aged ≥60 from the data of the pilot trial for secondary analysis.

2.2. Measurements

The PHQ-9 consists of nine items, each evaluating one of the nine DSM-IV criteria of MDD in the past two weeks. Each of the nine items was rated on a 4-point scale, ranging from 0 (not at all) to 3 (nearly every day)—summing up to a total PHQ-9 score, with higher scores indicating increased likelihood for MDD. Consistent with the process for translation of survey instrument outlined in prior studies[11], translation and back translation of the Chinese version of the PHQ-9 were repeated until the translators thought the Chinese version corresponded closely with the English version. The Chinese version of the PHQ-9 has been validated for primary care adults and adolescent students in Taiwan[9,12].

The Schedule for Clinical Assessments in Neuropsychiatry (SCAN) is a set of clinical instruments aiming to assess psychopathology and behavior associated with major psychiatric disorders in adults. The mood module of the SCAN was our criterion standard measure for MDD[13]. There was good interrater reliability between a research psychiatrist and the research nurses (generalized kappa of 0.88 for diagnosis of MDD according to DSM-IV standards) in the current study.

Measurement of depression severity was performed with the 17-item of Hamilton Rating Scale for Depression (HAMD-17), a widely used interview-based instrument to assess the severity of MDD. Each question has between 3 and 5 possible responses that increase in severity. The Chinese version of the HAMD-17 has been demonstrated to have good interrater reliability and internal reliability[14].

The Short Form of Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q-SF) uses the 14 general activities of the regular form (Q-LES-Q) as well as two global items with higher scores indicating better enjoyment and satisfaction with life. The scoring of Q-LES-Q-SF involves summing only the first 14 items to yield a total score. We used Chinese version of Q-LES-Q-SF as the measure for quality of life. The internal consistency of the Q-LES-Q-SF was 0.87[15].

2.3. Study procedures

After the participants agreed to participate and signed the informed consent, they were asked to complete the PHQ-9 and the Q-LES-Q-SF. Then, the trained research staff who was blinded to the results from the PHQ-9 interviewed them using the depression module of the SCAN and the HAMD-17. A subsample of participants was invited to re-take the PHQ-9 again within two weeks. All instruments were given on the same day in the context of their clinic appointment.

2.4. Statistical analyses

Data were analyzed with SPSS 21.0 software (IBM, Armonk, NY, USA). MDD and non-MDD patients were compared on demographic characteristics and mean scores of each measure (*t* test, chi-square, fisher's exact test, and adjust *p*-values for multiple comparisons with Bonferroni method). Inter-item reliability of the PHQ-9 was established by calculating Cronbach's coefficient α for the nine-item scale. Test-retest reliability within two-week interval was examined by using the intraclass correlation coefficient (ICC). The concurrent validity was examined by calculation of the Pearson's correlation coefficient (r) between the PHQ-9 and the appropriate instruments.

The exploratory factor analysis (EFA) was conducted with principal component analysis and varimax rotation to evaluate the construct validity. The confirmatory factor analysis (CFA) was performed with SPSS AMOS 18.0 software (IBM). The chi-square (χ^2) test, degrees of freedom (df), goodness-of-fit index (GFI), adjusted GFI (AGFI), and the root of the mean square residual (RMR) were applied to assess the fit of the model. Criterion validity tested the performance of PHQ-9 and its two subscales in comparison to the criterion standard, using SCAN diagnosis. Receiver-operating characteristic (ROC) curves were generated to compare the ability of these screening instruments, namely, PHQ-9, PHQ-2, and PHQ-1. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated. The optimal cutoff point for each instrument was determined by ROC analysis with the highest sum of sensitivity and specificity. All data analyses used two-tailed tests; $p < 0.05$ was considered significant differences.

3. Results

3.1. Participant characteristics

A total of 799 patients were approached and 634 subjects completed the assessments. Most of them were from community-based primary care settings (89.7%, n = 569), and the rest was from hospital-based family physician clinics (10.3%, n = 65). The participants were predominant female (58.8%) and low education levels. One-month prevalence rate of MDD was 2.2%. Characteristics of the sample according to MDD status were showed in Table 1. The mean score of PHQ-9 for full sample were 2.6 (SD = 3.7). Compared with non-MDD individuals, those with MDD were more significantly likely to be female, the separated/divorced, and have low socio-economic status (SES).

Please cite this article in press as: Chen I-P, et al., Validation of the Patient Health Questionnaire for Depression Screening Among the Elderly Patients in Taiwan, International Journal of Gerontology (2016), http://dx.doi.org/10.1016/j.ijge.2016.05.002