General Hospital Psychiatry 35 (2013) 592–597

Contents lists available at ScienceDirect

General Hospital Psychiatry

journal homepage: http://www.ghpjournal.com

Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: a cross-sectional study☆,☆☆,★

Masatoshi Inagaki, M.D., Ph.D. [a,b,*], Tsuyuka Ohtsuki, Ph.D. [b], Naohiro Yonemoto, Ph.D. [c], Yoshitaka Kawashima, Ph.D. [d], Akiyoshi Saitoh, Ph.D. [d], Yuetsu Oikawa, M.D. [e], Mie Kurosawa, M.D., Ph.D. [f], Kumiko Muramatsu, M.D., Ph.D. [g], Toshi A. Furukawa, M.D., Ph.D. [h,i], Mitsuhiko Yamada, M.D., Ph.D. [d]

[a] Department of Neuropsychiatry, Okayama University Hospital, Okayama, Japan
[b] Center for Suicide Prevention, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan
[c] Translational Medical Center, National Center of Neurology and Psychiatry, Tokyo, Japan
[d] Department of Neuropsychopharmacology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan
[e] Oshu City Magokoro Hospital, Iwate, Japan
[f] Iwate Mental Health Center, Iwate, Japan
[g] Clinical Psychology Course, Graduate School of Niigata Seiryo University, Niigata, Japan
[h] Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan
[i] Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan

ARTICLE INFO

Article history:
Received 17 May 2013
Revised 22 July 2013
Accepted 6 August 2013

Keywords:
Depression
Primary care
Internal medicine
Screening
Patient Health Questionnaire

ABSTRACT

Objective: Two depression screening tools, Patient Health Questionnaire (PHQ)-9 and PHQ-2, have not had their validity examined in general internal medicine settings in Japan. We examined the validity of these screening tools.
Methods: A total of 598 outpatients of an internal medicine clinic in a rural general hospital were enrolled consecutively and stratified by PHQ-9 score. Seventy-five patients randomly selected and 29 patients whose results from the PHQ-9 were considered to be positive for depressive disorder were then interviewed with a semistructured interview, the Mini International Neuropsychiatric Interview. We calculated diagnostic accuracy of the PHQ-9 and PHQ-2 to detect major depression and that of the suicidality item of the PHQ-9 to detect suicidality using sampling weights with multiple imputations.
Results: Sensitivity and specificity for depression were 0.86 and 0.85, respectively, for the PHQ-9 with cutoff points of 4/5, and 0.77 and 0.95, respectively, for the PHQ-2 with cutoff points of 2/3. Sensitivity and specificity of the suicidality item of the PHQ-9 were 0.70 and 0.97, respectively.
Conclusion: In internal medicine clinics in Japanese rural hospitals, the PHQ-2 with an optimal cutoff point for each setting plus the suicidality item of the PHQ-9 can be recommended to detect depression without missing suicidality.

1. Background

Depression is prevalent [1–4], and its influence on quality of life is profound. In middle-income and high-income countries including Japan, depression was the leading cause of disability in 2004 [5]. Prevalence of depression is reported to be high in health care settings [6–8]. The World Health Organization performed a primary care mental health survey across 14 countries and found that 14% of primary care patients suffered from major depression [7]. However, a meta-analysis showed that many depressed patients were overlooked and that many nondepressed patients were misdiagnosed as having depression [9]. Thus, development of a valid screening tool is required in health care settings.

The Patient Health Questionnaire (PHQ)-9 is one of the commonly used validated screening tools for depression in various settings including primary care and general practice [10–14]. The PHQ-9 has

☆ Authors' contributions: All authors have read and approved the final version of the manuscript. M.I. was the principal investigator and developed the original idea for the study. M.I., T.O., N.Y., Y.O., T.A.F. and M.Y. designed the study. M.I., T.O., N.Y. and T.A.F. analyzed data, and all authors discussed and prepared the manuscript.

☆☆ Funding: This work was supported by a grant for Research on Psychiatric and Neurological Disease and Mental Health from the Ministry of Health, Labour and Welfare. We thank Ms. Kyoko Sakurai, Ms. Hiromi Muramatsu and Ms. Mayumi Matsutani for their kind support.

★ The authors declare that they have no competing interests.

* Corresponding author. Department of Neuropsychiatry, Okayama University Hospital, 2-5-1, Sikada-cho, Kita-ku, Okayama, 700-8558, Japan. Tel.: +81 86 235 7242; fax: +81 86 235 7246.
E-mail addresses: masatoshiinagaki@okayama-u.ac.jp (M. Inagaki), ohtsuki@ncnp.go.jp (T. Ohtsuki), yonemoto@ncnp.go.jp (N. Yonemoto), kawashima@ncnp.go.jp (Y. Kawashima), asaitoh@ncnp.go.jp (A. Saitoh), yuetsu-o@city.oshu.iwate.jp (Y. Oikawa), miek@pref.iwate.jp (M. Kurosawa), muramatu@n-seiryo.ac.jp (K. Muramatsu), furukawa@kuhp.kyoto-u.ac.jp (T.A. Furukawa), mitsu@ncnp.go.jp (M. Yamada).

0163-8343/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.genhosppsych.2013.08.001

two different assessment methods. One uses a categorical (yes/no) algorithm to screen for depression. In the categorical algorithm, the depression screening is positive if five or more of the nine depressive symptom criteria are endorsed and one of the symptoms is anhedonia or depressed mood. The other is a dimensional (continuous scale) assessment. In the dimensional assessment, the sum of the PHQ-9 scores is calculated, and an a priori cutoff score is applied. Using the categorical algorithm, the sensitivity and specificity shown in a meta-analysis were 0.77 [95% confidence interval (CI): 0.71–0.84] and 0.94 (95% CI: 0.90–0.97), respectively [14]. In comparison, the sensitivity and specificity of the dimensional assessment using cutoff scores for the PHQ-9 in primary care settings that were calculated in another meta-analysis were 0.86 (95% CI: 0.66–0.97) and 0.88 (95% CI: 0.80–0.93), respectively [12]. A third meta-analysis showed the sensitivity and specificity of the dimensional assessment using the cutoff score 9/10 as 0.81 (95% CI: 0.72–0.88) and 0.92 (95% CI: 0.83–0.97), respectively [10]. However, these studies showed major variability in the optimal cutoff point [10,12]. A cutoff point of around 10 was recommended in a previous publication [15] and shown to be the optimal cutoff point in a previous study using meta-analysis [12]. However, the optimal cutoff point score has been reported as lower in elderly populations [16,17], a population in Korea [16] and a community sample in rural Pakistan [18]. In addition, the sensitivity using this cutoff point was relatively low (sensitivity: 0.50–0.69) [19–21] in some settings including a hospital setting [22,23] and a family practice clinic (sensitivity: 0.53) [24].

In Japan, especially in rural areas, most patients consult a general internist who plays a role similar to that of a primary care physician or a general practitioner. We reported the prevalence of depression in an internal medicine outpatient clinic in a rural general hospital as 7.4% [25]. In that study, we showed that a critical symptom, suicidality, was present in 12.7% [25]. In a previous study, however, we also reported that physicians overlooked many symptoms of depression and did not pay attention to diagnosing depression [26]. Thus, a screening tool to support depression identification has to be developed and used widely. However, there is no screening tool for which the validity has been confirmed in such primary care settings in Japan.

Therefore, in the present study, we examined the validity of the Japanese versions of the PHQ-9 and PHQ-2 in an internal medicine outpatient clinic in a rural general hospital, which has a role in the primary care of the residents in the area. In addition, we examined the validity of the suicidality item of the PHQ-9 to detect suicide ideation as ascertained by a semistructured interview because the PHQ-2 does not include the question of suicidality and may cause the clinician to miss this critical symptom.

2. Methods

2.1. Participants

The sampling process and procedures of the study have been reported previously [25] and are described briefly below. This study was conducted on 9 consecutive consultation days between July 12 and 23, 2010, at an internal medicine outpatient clinic in a general hospital having no mental health specialties. This hospital is located in a small city (population of 124,756 in 2010) in the Tohoku region of Japan. The hospital serves as a regional public hospital and is funded by the National Health Insurance Society in Oshu. The city is located in a typical rural area about 500 km north of Tokyo with low population influx. There are high proportions of elderly people and people engaged in primary industry [27].

We used the following inclusion criteria to define a target population to be assessed for depression: (1) patients aged 20 or older who visited the outpatient clinic to consult a physician for their own primary care and (2) patients who have no communication difficulties, such as hearing loss or language problems, and who have no severe cognitive impairment, such as dementia or disturbance of consciousness. Thus, we did not include visitors who came in for admission preparation or those who consulted for their family members. We did not include patients who lived outside the catchment area of the hospital. Severe cognitive impairment was judged based on a semistructured interview using the first two questions of the Mini-Mental State Examination concerning time and place orientation [28,29]. This was administered by research staff consisting of two psychiatrists (M.I. and M.Y.), a research assistant (T.O.) having experience in surveys using the Mini-Mental State Examination and PHQ-9 in internal medical clinics, and nurses. The staff sometimes conducted an additional interview about patients' lifestyle factors and dementia history if accompanying persons were present. For ethical reasons and for the feasibility of the survey, we also excluded patients who were too physically ill to be interviewed.

This study was approved by the ethics committee of the National Center of Neurology and Psychiatry in Japan. The researchers provided all participants with detailed information using a written document and administered a battery of self-report questionnaires (PHQ-9) [15,30] after the patients provided oral informed consent. After this first-stage screening to stratify the participants, we conducted the structured psychiatric interviews [Mini International Neuropsychiatric Interview (MINI)] [31,32] with patients who provided further written informed consent.

2.2. Measurements

2.2.1. PHQ-9

The PHQ-9 is a widely used screening tool in health care settings [15,30]. The PHQ-9 is a self-report questionnaire consisting of nine questions asking about depression symptoms such as anhedonia, depressed mood and suicidality. The PHQ-9 was translated into Japanese (the Japanese version of PHQ-9) through back translation and validated in a previous study [30]. We asked patients to choose from the following options how often they had been bothered by each of the nine symptoms over the last 2 weeks: "Not at all," "Several days," "More than half the days" and "Nearly every day."

A categorical algorithm has been proposed in the literature [15,30] to assess the PHQ-9 results. In the categorical algorithm, the depression screening is positive if five or more of the nine depressive symptom criteria are endorsed as having been present at least "more than half the days" and one of the symptoms was anhedonia or depressed mood. One of the nine items, "thoughts that you would be better off dead or of hurting yourself in some way," is counted if present at all. In addition to the categorical algorithm of the PHQ-9, we adopted a two-item categorical algorithm with the first two items of the PHQ-9 (anhedonia and depressed mood). In the two-item categorical algorithm, depression screening is positive if one or more of the two depressive symptom criteria are endorsed as having been present.

Alternatively, we calculated the sum of the PHQ-9 scores as a dimensional assessment of the PHQ-9. Each item of the PHQ-9 is scored from 0 to 3, with a total possible score of 27 for the nine items. Also, we calculated the sum of the PHQ-2. A total possible score of the PHQ-2 was 6.

2.2.2. MINI

We diagnosed major depressive episode using the Major Depression Episode module of the MINI [31,32]. The interview was originally developed as a semistructured diagnostic interview compatible with the *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition,* (DSM-III-R) and *International Classification of Diseases, 10th Revision,* criteria [33,34]. The MINI was translated in a previous study [31] through the standard procedure of back translation for the cross-cultural adaptation of an original English psychometric instrument [35]. The Japanese version of the MINI was validated using the