Journal of Affective Disorders 170 (2015) 138–142

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany

Juliana J. Petersen [a,*], Michael A. Paulitsch [a], Johannes Hartig [b], Karola Mergenthal [a], Ferdinand M. Gerlach [a], Jochen Gensichen [a,c]

[a] Institute of General Practice, Goethe-University Frankfurt am Main, Frankfurt am Main, Germany
[b] Department of Educational Quality and Evaluation, German Institute for International Educational Research, Frankfurt am Main, Germany
[c] Institute of General Practice, Friedrich Schiller University Hospital Jena, Jena, Germany

ARTICLE INFO

Article history:
Received 20 April 2014
Received in revised form
28 August 2014
Accepted 29 August 2014
Available online 6 September 2014

Keywords:
Primary healthcare
Major depression
Psychiatric status rating scales (MESH terms)

ABSTRACT

Background: Depression is characterized by gender-specific distinctions, with women being affected more often than men. The Patient Health Questionnaire-9 (PHQ-9) is frequently used to assess depression in primary healthcare. Previous research has yielded heterogeneous findings on the factor structure, and little is known of its measurement invariance across gender. The aim of this study was 1) to evaluate the fit of four previously hypothesized models of PHQ-9 factor structure in patients with major depression in German family practices, and 2) to test the measurement invariance of the best-fitting model across gender.
Methods: We used the baseline data from a cluster-randomized controlled trial. The diagnosis of major depression was based on the PHQ-9 and confirmed by the family physician. We calculated Confirmatory Factor Analyses (CFA) to assess which of the previously hypothesized factor structures (a one- and three different two-factor solutions) would best fit our data. We also calculated Complex Survey Analyses (CSA) and Multi Sample Analyses (MSA).
Results: We included 626 participants (75.4% women and 24.6% men). A two-factor model with five 'somatic' labeled items and four 'non-somatic' labeled items presented the best fit indices. The model measurement was invariant across gender.
Limitations: The inclusion criteria used in the main trial mean the study sample was not representative of all patients with major depression in German family practices.
Conclusions: The measurement invariance across gender revealed by this study is a precondition for the use of the PHQ-9 without gender-specific adaptation in patients with major depression in German family practices.

1. Introduction

Depression is a leading cause of disease burden (Whiteford et al., 2013) and characterized by gender-specific differences. Across cultures, the prevalence of major depressive disorder (MDD) is higher among women than men (Maier et al., 1999; Weissman et al., 1996), and gender differences in the clinical presentation of MDD have also been reported (Marcus et al., 2005; Schuch et al., 2014; Wenzel et al., 2005). Schuch et al., for instance, found that female patients are more likely to suffer from increased weight and somatic complaints (Schuch et al., 2014).

The Primary Care Evaluation of Mental Disorders (PRIME-MD) instrument was developed to assist primary care physicians in making criteria-based diagnoses of the five most common types of mental disorders presenting in medical populations, including depression. The Patient Health Questionnaire (PHQ) is the self-report version of the PRIME-MD that was developed to reduce the clinician time required to evaluate a patient's responses (Spitzer et al., 1999). The Patient Health Questionnaire Depression Scale (PHQ-9) is a 9-item depression module derived from the full PHQ. The items come directly from the nine signs and symptoms of

* Correspondence to:Institute of General Practice, Goethe-University Frankfurt am Main; Theodor-Stern-Kai 7; D-60590 Frankfurt/Main, Germany. Tel.: +49 69 6301 83883; Fax: +49 69 6301 6014.
E-mail address: petersen@allgemeinmedizin.uni-frankfurt.de (J.J. Petersen).

http://dx.doi.org/10.1016/j.jad.2014.08.053
0165-0327/

major depression delineated in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (Kroenke et al., 2001). The PHQ-9 has been frequently used to assess depression in primary healthcare, where the majority of patients are treated (Kroenke et al., 2010). Previous research has shown good validity (with a sensitivity of 0.77–0.88 and specificity of 0.88–0.94 for detecting depressive disorders) and reliability (Gilbody et al., 2007; Kroenke et al., 2010; Wittkampf et al., 2007). With regard to the underlying factor structure, however, previous research has yielded heterogeneous findings. While some studies found evidence of a one-factor structure (Baas et al., 2011; Cameron et al., 2008;Huang et al., 2006; Mewes et al., 2010; Yu et al., 2012), others supported slightly different two-factor models (de Jonge et al., 2007; Krause et al., 2008; Krause et al., 2010; Richardson and Richards, 2008). The two-factor models have in common that one factor is basically represented by 'somatic' items (e.g., sleep disturbances, fatigue and appetite changes) and the other factor by 'non-somatic' or 'affective' items (e.g., depressed mood, lack of interest, feelings of worthlessness and suicidal ideation). Previous research on the PHQ-9′s factor structure was performed in different populations, including patients with spinal cord injury (Kalpakjian et al., 2009; Krause et al., 2008; Krause et al., 2010; Richardson and Richards, 2008), patients with coronary heart disease (de Jonge et al., 2007), the general population (Yu et al., 2012), and primary care patients (Baas et al., 2011; Huang et al., 2006). Elhai et al. (2012) compared different models of previously hypothesized factor structures of the PHQ-9 by using Confirmatory Factor Analysis (CFA) in a sample of soldiers, and found that their data best fitted a two-factor model with five somatic and four non-somatic symptoms. These results supported studies conducted by Krause et al. (2010) and Richardson and Richards (2008) in patients with spinal cord injury.

Despite pronounced gender-specific differences in depression and the frequent use of the PHQ-9 in primary care, little research has been conducted into whether the factor structure differs between male and female patients. This is an important issue, since a precondition for the comparability of results is that the instrument assesses the same latent construct (depression) in men as in women. Yu et al. (2012) used the PHQ-9 in a Chinese community sample in Hong-Kong and described a similar one-factor structure for both genders. Kalpakjian et al. (2009) carried out an exploratory factor analysis of the PHQ-9 in patients with spinal cord injury and found that the two-factor structure differed for men and women. Baas et al. (2011), who analyzed the factor structure of the PHQ-9 in primary care patients with a high risk for depression, found that it was measurement invariant for ethnicity in women and partially measurement invariant for ethnicity in men. A major limitation of the study was, however, that a gold standard measure of depression was not included.

The aim of this study was 1) to assess which of the previously hypothesized models representing alternative factor structures of the PHQ-9 would best fit the observed data in patients with major depression in German family practices and 2) to test the measurement invariance of the best-fitting model across gender.

2. Methods

2.1. Study design and sample

For this cross-sectional study we used baseline data from a cluster-randomized controlled trial conducted between 2005 and 2008 on the effectiveness of a collaborative care intervention for patients with major depression in German family practices (Gensichen et al., 2009). The inclusion criteria for patients in the trial were diagnosis of major depression with indication for an antidepressive treatment, aged 18–80, access to a private telephone, ability to give informed consent and to communicate in German. The diagnosis of major depression was based on a categorical diagnosis and a score of more than 9 points in the PHQ-9. Major depressive disorder should be considered in patients who indicate that five or more of these nine criteria were present on *more than half the days* during the previous two weeks and when one of the symptoms is depressed mood or anhedonia. One of the nine criteria ("Thoughts that you would be better off dead or of hurting yourself in some ways") counts regardless of duration. The diagnosis was confirmed by the GP by using the checklists in the *Diagnostic and Statistical Manual of Mental Disorders*, *Fourth Edition*. Exclusion criteria were confirmed pregnancy, alcohol abuse or illicit drug use and acute suicidal ideation, as assessed by the family physician.

2.2. Procedure and instruments

Data collection occurred by means of self-rating questionnaires for patients and data collection from patient records. We employed the validated German version of the PHQ-9 to assess depression severity (Löwe et al., 2002). Patients were asked to select the frequency of depressive symptoms they had experienced in the past 2 weeks on a Likert-scale (0 'not at all', 1 'several days', 2 'more than half the days', and 3 'nearly every day'). The PHQ-9 comprises the following 9 items: 1) little interest or pleasure in doing things, 2) feeling down, depressed, or hopeless, 3) trouble falling or staying asleep or sleeping too much, 4) feeling tired or having little energy, 5) poor appetite or overeating, 6) feeling bad about yourself – or that you are a failure or have let yourself or your family down, 7) trouble concentrating on things, such as reading the newspaper or watching television 8) moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual and 9) thoughts that you would be better off dead or of hurting yourself in some way (a full PHQ-9 version is available online at http://www.phqscreeners.com). The PHQ-9 can be used either as a continuous measure or as a diagnostic algorithm to make a probable diagnosis of major depressive disorder. When used as continuous measure, the total score can be calculated by summing up the responses, and thus ranges from 0 to 27 (higher scores indicate more severe depression) (Kroenke et al., 2001).

We used written consent procedures for family physicians and patients, and all participants gave informed consent. The institutional review board of Goethe-University Frankfurt am Main approved the study protocol.

2.3. Previously hypothesized factor structures of the PHQ-9

Several studies have used principal component analysis (PCA) (e.g., (Cameron et al., 2008; Huang et al., 2006)) and exploratory factor analysis (EFA) (e.g., (Mewes et al., 2010; Richardson and Richards, 2008)) as dimension-reducing procedures to identify a small set of synthetic variables, called eigenvectors or factors, that explain most of the total (principal component analysis) or common (exploratory factor analysis) variation present in the PHQ-9 items. These techniques represent a good means for exploring a latent factor structure, in case no theoretical or empirical hypothesis on the factor structure is available (Bryant and Yarnold, 1995). Other studies have used confirmatory factor analysis (CFA) to test hypothesized models generated from EFA studies (e.g., (Baas et al., 2011;Elhai et al., 2012; Yu et al., 2012)).

For this study we chose to analyze four of the most commonly described models derived from PCA/EFA or CFA studies. The first model ('Model 1') posits that a single one-dimensional factor underlies the PHQ-9. This model is based on PCA/EFA studies that found support for a one-factor PHQ-9 model in samples of primary