Psychiatry Research 199 (2012) 169–173

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

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

The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9

Jon D. Elhai[a,b,*,1], Ateka A. Contractor[a], Marijo Tamburrino[b], Thomas H. Fine[b], Marta R. Prescott[c], Edwin Shirley[d,e], Phillip K. Chan[d,e], Renee Slembarski[d,e], Israel Liberzon[f,g], Sandro Galea[c], Joseph R. Calabrese[d]

[a] Department of Psychology, University of Toledo, Toledo, OH, USA
[b] Department of Psychiatry, University of Toledo, Toledo, OH, USA
[c] Mailman School of Public Health, Columbia University, New York, NY, USA
[d] Department of Psychiatry, Case Western Reserve University, Cleveland, OH, USA
[e] University Hospitals Case Medical Center, Cleveland, OH, USA
[f] Mental Health Service, VA Ann Arbor Health System, Ann Arbor, MI, USA
[g] Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Article history:
Received 21 September 2011
Received in revised form
10 December 2011
Accepted 20 May 2012

Keywords:
Depression
Confirmatory factor analysis
Military veterans
Patient Health Questionnaire-9

ABSTRACT

Little research has examined the underlying symptom structure of major depressive disorder (MDD) symptoms based on *DSM-IV* criteria. Our aim was to analyze the symptom structure of major depression, using the Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 was administered to a sample of 2615 Army National Guard soldiers from Ohio. A one-factor model of depression and three separate two-factor models previously established in the literature were evaluated using confirmatory factor analysis. Results demonstrated greater support for the two-factor models of depression than for the one-factor model. The best fitting model was the two-factor model of somatic and non-somatic symptoms supported previously by Krause et al. (2010) and Richardson and Richards (2008). Implications for understanding the components and mechanisms of MDD are discussed.

1. Introduction

A substantial body of research has investigated the underlying symptom structure of depression symptoms (Shafer, 2006). However, this body of research has rarely used depression measures with items that map onto *DSM-IV* symptom criteria for major depressive disorder (MDD). Thus, little is known about the symptom structure of *DSM-IV*-based MDD. The Patient Health Questionnaire-9 (PHQ-9), on the other hand, is one of the few depression screening instruments comprising items that map onto *DSM-IV* diagnostic criteria for MDD symptoms. Although the PHQ-9 has been well-researched and used, few studies have assessed its factor structure.

The PHQ-9 is a nine-item major depression module from the Primary Care Evaluation of Mental Disorders (PRIME-MD), a standardized assessment for mood, anxiety, somatoform, alcohol-related, and eating disorders (Spitzer et al., 1994). It was originally designed for the purpose of screening in primary care, an important issue because primary care is the most frequent setting in which individuals seek treatment for mental health reasons (Wang et al., 2006). The PHQ-9 items are self-rated and map onto *DSM-IV* symptom criteria for MDD. An additional item assesses the impact of depressive symptoms on functional impairment. Psychometric evidence demonstrates adequate construct and criterion validity across diverse samples of civilian primary care and obstetrics–gynecology patients (Kroenke et al., 2010; Flynn et al., 2011). Total cut-off scores of 5, 10, 15 and 20 represent "mild," "moderate," "moderately severe," and "severe" (clinically diagnosable) depression respectively; although these labels should not be confused with similar specifiers for major depression from *DSM-IV*. The PHQ-9 diagnostic scoring algorithm corresponds to *DSM-IV* MDD requirements, wherein five or more items must be recently experienced at least "more than half the days," with one symptom being depressed mood and/or anhedonia (Kroenke et al., 2001).

There is a paucity of research on the factor structure of *DSM-IV*-based MDD symptoms. The PHQ-9 is an instrument that queries such symptoms. It is true that the Major Depression

* Corresponding author at: Department of Psychology, University of Toledo, Mail Stop #948, 2801 W. Bancroft Street, Toledo, OH 43606-3390, USA.
Tel.: +1 419 530 2829; fax: +1 419 530 8479.
E-mail address:
[1] Correspondence should be initiated at authors website www.jon-elhai.com

0165-1781/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psychres.2012.05.018

Inventory (MDI) (Olsen et al., 2003) is another instrument of *DSM-IV*-based MDD symptoms. Despite being analyzed using item response theory which assumes unidimensionality, the MDI's factor structure has not been investigated. Research is necessary in understanding the latent structure and mechanisms underlying MDD, and studying the PHQ-9 is one means of examining this relatively unexplored research area.

Several studies have used exploratory factor analysis (EFA) to discover relatively unique sets or "factors" of PHQ-9 items that are correlated with each other, in samples of primary care patients, outpatient substance abusers and spinal cord injury patients (Huang et al., 2006; Cameron et al., 2008; Dum et al., 2008; Kalpakjian et al., 2009; Krause et al., 2010). Other studies have used confirmatory factor analysis to test hypothesized models generated from EFA studies, using samples of spinal cord injury and primary care patients (Krause et al., 2008; Baas et al., 2011). Some EFA studies (Huang et al., 2006; Cameron et al., 2008; Dum et al., 2008; Kalpakjian et al., 2009) and CFA studies of the PHQ found support for a one-factor model (Baas et al., 2011), reflecting unidimensionality of the depression construct. Other EFA studies (Richardson and Richards, 2008; Krause et al., 2010) and a CFA study (Krause et al., 2008) found support for a two-factor PHQ model, with one factor primarily based on somatic items (e.g., sleep difficulties, appetite changes and fatigue) and the other factor primarily based on non-somatic or affective items (e.g., depressed mood, feelings of worthlessness and suicidal thoughts). Importantly, these studies typically investigated only one depression model per paper. Lacking in the literature is a comprehensive evaluation of the various depression factor models, from a *DSM-IV*-based instrument such as the PHQ-9, tested against each other using objective statistical criteria.

We believe that using CFA is more defensible in examining the structure of MDD symptoms, and we built on this area in the present paper. EFA is exploratory and capitalizes on chance error in statistically "discovering" the best patterns of item sets based on computer-generated mathematical algorithms. Given the substantial Type I error, and lack of ability to narrowly specify hypothesized models of interest in EFA a priori, EFA findings do not often cross-validate with other samples (Fabrigar et al., 1999). Unlike EFA, CFA requires the testing of a specifically hypothesized model (or models), evaluating how well the model's patterns of intercorrelations are represented in the observed data (Bollen, 1989; Kline, 2010).

The PHQ-9's depression models derived from EFA and CFA have important implications for understanding the core components of depression—specifically, with implications for underlying dimensions and mechanisms of the disorder. For example, three PHQ-9 items have consistently been found represented by the somatic factor (sleep difficulties, appetite changes and fatigue); three additional items have been represented by the affective/non-somatic factor (depressed mood, feelings of worthlessness and thoughts of death) (Krause et al., 2008, 2010; Richardson and Richards, 2008). These preliminary studies thus suggest somatic and affective components of MDD that may have implications for the disorder's underlying mechanisms. Additionally, the psychomotor difficulties item was less represented by the somatic and non-somatic factors (Krause et al., 2008, 2010), thus possibly reflecting that this symptom may not be as strong of a general depression marker.

The current study aimed to study the factor structure of *DSM-IV*-based MDD symptoms by comparing four empirically-supported models using CFA to assess which model fits best. We used the PHQ-9 to examine our research question. We used objective statistical criteria to judge and compare model fit. Our sample included a large, regional epidemiological sample of National Guard soldiers, many of whom had been deployed to a war zone which places soldiers at risk of depression soon and months after returning from combat (Grieger et al., 2006). This sampling plan broadens previous research on the PHQ-9's factor structure from the restricted samples that have included patients with spinal cord injury (Krause et al., 2008, 2010; Richardson and Richards, 2008; Kalpakjian et al., 2009), and primary care patients (Huang et al., 2006; Baas et al., 2011). We found no studies examining the PHQ's factor structure among a military population.

2. Method

2.1. Participants and procedure

The present study was part of the Ohio Army National Guard Mental Health Initiative, a large-scale epidemiological study of mental health among National Guard soldiers in Ohio. We initially invited all members of the Ohio National Guard who served between July 2008 and February 2009 for participation in the telephone interview portion of the study, executed through an alert letter. 12,225 Guard members with a valid mailing address were invited to participate (345 additional individuals were excluded for having no mailing address). Among the pool of potential subjects, 1013 (8.3%) declined to participate, 1130 (10.1%) did not have a telephone number listed with the Guard, and 3568 (31.8%) did not have a correct or working phone number. Among the remaining 6514 Guard members (58.1%), the following individuals were excluded: 187 (2.8%) based on age eligibility restrictions, 1364 (20.9%) declined to participate, 31 (0.4%) for having English language or hearing difficulties, and 2316 (35.5%) for not being contacted before the cohort was closed to new recruitment. Of the remaining 2616 subjects, one subject was excluded for missing six items on our primary measure (PHQ-9), leaving 2615 participants.

Age ranged from 17 to 61 years (M=30.69, S.D.=9.50). Most participants were men (n=2227, 85.2%), with 388 women (14.8%). The vast majority of participants were either currently married (n=1227, 47.0%) or had never been married (n=1133, 43.4%). A small proportion of respondents were of Hispanic ethnicity (n=82, 3.1%). Most identified their racial background as Caucasian (n=2295, 87.9%), or African American (n=194, 7.4%). The majority had some college or technical school education (n=1233, 47.2%), a completed college education (n=466, 17.8) or a terminal high school or equivalent degree (n=598, 22.9%). Participants were primarily employed full-time (n=1560, 59.9%), part-time (n=340, 13.1%), unemployed (n=419, 16.1%) or of student status (n=244, 9.4%). Household income of $20,000 or less was reported by 326 participants (12.9%), between $20,001 and $40,000 by 604 respondents (23.8%), between $40,001 and $60,000 by 568 participants (22.4%), between $60,001 and $80,000 by 426 (16.8%), and more than $80,000 by 611 participants (24.1%). Service in the military ranged from 0 to 50 years (M=10.08, S.D.=8.41). Roughly one-third of the sample had never been deployed (n=939, 36.0%), while most deployed participants reported being deployed once (n=816, 31.3%), twice (n=469, 18.0%), or three times (n=213, 8.2%). Exactly 765 participants (46.0%) had been deployed to Iraq or Afghanistan between 2003 and 2009.

Study enrollment began in November 2008 and ended in November 2009. The National Guard Bureau, Office of Human Research Protections of the U.S. Army Medical Research and Materiel Command, along with several affiliated hospital and university institutional review boards, approved the study, with written informed consent waived in lieu of verbal consent by telephone. We complied with the Declaration of Helsinki in ethically conducting this research.

2.2. Instrumentation

A computer-assisted telephone interview was conducted for all participants by trained professionals, to assess demographic characteristics and mental health functioning using standardized questionnaires.

Of relevance to the present study, the PHQ-9 was administered by telephone (Kroenke et al., 2001). The PHQ-9 measures depression symptoms over the previous 2 weeks based on *DSM-IV* major depressive episode symptom criteria. We modified the instructions in order to query depression symptoms over the course of the respondent's lifetime; support for this approach was found in Cannon et al. (2007). The PHQ-9 is a Likert-type self-report instrument with four response options ranging from "*0=Not at all*" to "*3=Nearly every day*." As a severity measure, scores on the PHQ-9 range from 0 to 27. Internal consistency has ranged from 0.86 to 0.89 (Kroenke et al., 2001) (alpha=0.86 in the present sample). Test–retest reliability within 48 h was r=0.84. Diagnostic validity has been demonstrated in detecting an MDD diagnosis based on structured diagnostic interviews. Lastly, construct validity is reflected in the association of PHQ-9 severity scores and measures of functional status, number of disability days and difficulties based on symptoms (Kroenke et al., 2001).

Table 1
Factor models tested, and factors on which items were mapped..

PHQ 9 Items	Model 1	Model 2a	Model 2b	Model 2c
1. Anhedonia	Depression	Non-somatic	Non-somatic	Somatic
2. Depressed mood	Depression	Non-somatic	Non-somatic	Non-somatic
3. Sleep difficulties	Depression	Somatic	Somatic	Somatic
4. Fatigue	Depression	Somatic	Somatic	Somatic
5. Appetite changes	Depression	Somatic	Somatic	Somatic
6. Feeling of worthlessness	Depression	Non-somatic	Non-somatic	Non-somatic
7. Concentration difficulties	Depression	Non-somatic	Somatic	Somatic
8. Psychomotor agitation/retardation	Depression	Non-somatic	Somatic	Somatic
9. Thoughts of death	Depression	Non-somatic	Non-somatic	Non-somatic

Note. Model 2a is based on Krause et al. (2010) EFA model at baseline measurement and Krause et al. (2008) CFA model. Model 2b is based on Krause et al. (2010) EFA model at 17 months post-injury, and Richardson and Richards' (2008) model at 1 year post-injury. Model 2c is based on Krause et al. (2010) EFA model at 29 months post-injury.

2.3. Analysis

Of the 2615 participants, 62 were missing between one to two PHQ items each, and one subject missed three items. We used multiple imputation procedures with an iterative Markov chain Monte Carlo method and the Gibbs Sampler procedure (in SPSS's Version 17 Missing Value Analysis software) to estimate missing item-level PHQ data, generated across 10 imputed datasets. CFA analyses were conducted using Mplus 6.1 software, averaging parameter estimates across the 10 imputed datasets. Three PHQ items had kurtosis values greater than 2.0, thus deemed non-normally distributed. Therefore CFA analyses implemented maximum likelihood estimation with a mean-adjustment, using the Satorra–Bentler χ^2 value which is robust to non-normality (Satorra and Bentler, 2001).

CFA analyses were implemented by specifying the four models discussed above (and displayed in Table 1), treating PHQ items as continuously-scaled. All residual error covariances were fixed to zero. In scaling the factors within a model, we fixed factor variances to 1. Goodness of fit indices are reported below, including the comparative fit index (CFI), Tucker Lewis Index (TLI), root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR). Models fitting very well (or adequately) are indicated by CFI and TLI $\geq$0.95 (0.90–0.94), RMSEA $\leq$06 (0.07–0.08), SRMR $\leq$0.08 (0.90–0.10) (Hu and Bentler, 1999). All tests were two-tailed. Comparing nested models by examining differences in traditional goodness of fit indices (mentioned above) is not appropriate, and inaccurate (Fan and Sivo, 2009). Therefore, in comparing the one-factor model with a given two-factor model, we used a difference test for nested models between Satorra–Bentler χ^2 values (with a correction factor because the Satorra–Bentler χ^2 value is not distributed normally on a χ^2 distribution) (Muthén and Muthén, 2006). We also present Bayesian Information Criterion (BIC) values for comparing the two-factor models with each other; χ^2 difference testing is not possible between the two-factor models since they are not nested within one another. In comparing BIC values between models, a 10-point BIC difference represents a 150:1 likelihood and "very strong" ($p<0.05$) support that the model with the smaller BIC value fits best; a difference in the 6–9 point range indicates "strong" support (Kass and Raftery, 1995; Raftery, 1995).

Table 2
Comparison of fit statistics for the four PHQ factor analytic models.

Model	Satorra–Bentler χ^2 (d.f.)	CFI	TLI	RMSEA	SRMR	BIC	Compared to Model 1 χ^2_{Diff} (d.f.)
1	317.71(27)*	0.94	0.91	0.06	0.04	56,476.07	
2a	246.62(26)*	0.95	0.93	0.06	0.04	56,356.37	69.64(1)*
2b	**210.35(26)***	**0.96**	**0.94**	**0.05**	**0.03**	**56,290.75**	**101.02(1)***
2c	261.63(26)*	0.95	0.93	0.06	0.04	56,382.56	53.87(1)*

Note: *CFI*=Comparative fit index; *TLI*=Tucker–Lewis Index; *RMSEA*=Root mean square error of approximation; *SRMR*=standardized root mean square residual; *BIC*=Bayesian information criterion.

* $p<0.001$.

Table 3
Standardized factor loadings for Model 2b.

Factor loadings	β
Non-somatic	
1. Anhedonia	0.76
2. Depressed mood	0.79
6. Feelings of worthlessness	0.68
9. Thoughts of death	0.48
Somatic	
3. Sleep difficulties	0.68
4. Fatigue	0.71
5. Appetite changes	0.65
7. Concentration difficulties	0.65
8. Psychomotor agitation/retardation	0.50

3. Results

We first estimated descriptive statistics for the PHQ-9 total score. The sample ranged in score from 0 to 27, with a mean of 5.12 (S.D.=6.01). Thus this sample, as a whole, was represented by only mild depression. Based on the diagnostic scoring algorithm described above and discussed by Kroenke et al. (2001), 282 participants (10.8%) would be classified as a "probable depression" case. Based on a more liberal screening method, using a cutoff score of 10 or higher, 564 participants (21.6%) would be classified as a "probable depression" case.

The one-factor model (Model 1) fit the data reasonably well (see Table 2), as indicated by relatively high estimates on CFI and TLI, and low estimates of RMSEA and SRMR. As demonstrated by χ^2 difference testing (implementing the Satorra–Bentler correction factor) each two-factor model fit significantly better than the one-factor model (with uniformly lower χ^2 values for the two-factor models); thus the PHQ-9 appeared to be multidimensional rather than unidimensional. Based on BIC values, Model 2b fit significantly better (with a smaller BIC value) than Model 2a, which in turn fit significantly better than Model 2c. Model 2b's superior fit is evidenced by a difference in BIC values of 66 when compared to Model 2a, and a difference of 92 when compared to Model 2c.[1]

Table 3 presents factor loadings for Model 2b, which was the best fitting model. All factor loadings were 0.65 or higher, except for the last loading per factor. Thus for the most part, the items within a given factor hung together very well. The correlation between factors was 0.87.

[1] Because CFA results sometimes do not replicate with other samples, we drew a random subsample of ~50% of participants (n=1304) and re-computed the CFA analyses, comparing those results with results computed for the remaining 1311 subjects. We found nearly identical results across subsamples and when compared to the results presented above, with only very minor discrepancies in fit indices. Furthermore, as in the findings presented above, in both randomly generated subsamples all fit indices were strongest for Model 2*b*.

4. Discussion

We empirically tested four *DSM-IV*-based depression models previously investigated using EFA or CFA in order to reveal the best fitting depression model. Our sample of Army National Guard soldiers was not a severely depressed sample; rather, they were represented as a whole by previously having only mild depression symptoms. As such, our tests of the latent structure of the PHQ-9's depression symptoms essentially modeled depression's components at the mild end of the depression continuum.

We found that using the PHQ-9, the two-factor model supported by Krause et al. (2010) using EFA from their study's second wave of data (17-month) of three longitudinal measurement waves, and by Richardson and Richards (2008), best fit the data. This model is different from the one-factor model previously supported in EFA studies because it separates the general depression factor into somatic and non-somatic factors. It differs from Krause et al.'s model found to fit best in the authors' first of three longitudinal measurement phases because the current model conceptualizes concentration and psychomotor difficulties as part of the somatic factor. The current model differs from Krause et al.'s model discovered from EFA at their study's third measurement wave (29-month follow up) because the current model conceptualizes anhedonia as part of the non-somatic factor.

Each of the two-factor models fit significantly better than the one-factor model. Thus, we find that MDD symptoms (using the PHQ-9) are best represented by somatic and non-somatic factors, rather than a single, unidimensional factor. This distinction between somatic and non-somatic factors is not only an internal distinction within a depression instrument (the PHQ-9), but likely represents differential relations with other types of psychopathology. For example, Elhai et al. (2011) found that using the Center for Epidemiologic Studies-Depression Scale with Canadian military veterans, depression's somatic items were significantly more related than depression's affective items to particular posttraumatic stress disorder (PTSD) factors. Furthermore, in the present study, the best fitting model (2b) fit better than the other two-factor models, determined by objective statistical criteria. However, because the present study is the first to use CFA with a *DSM-IV*-based depression instrument to test competing depression models, future studies are needed to further examine these, and perhaps other, competing models.

This research is important in contributing to an understanding of the true components of MDD. Future research should further explore the somatic and affective dimensions of MDD and evaluate which of these constructs is more related to depression-related psychopathology as well as treatment outcomes. Understanding these constructs and their correlates can potentially aid clinicians in prioritizing treatment interventions aimed at reducing the more severe subtype of these MDD symptoms. For example, if one subtype of MDD symptoms (for example, somatic items) is found more related to psychopathology, then perhaps that subtype would represent a priority for clinical intervention over other depression symptoms. Furthermore, MDD is similar to, overlaps with, and is comorbid with other mood and anxiety disorders. This feature of MDD will not likely cease with *DSM-5*'s publication, since MDD's proposal for *DSM-5* is quite similar to the *DSM-IV* version, as is the case for other mood and anxiety disorders with which MDD overlaps (American Psychiatric Association. DSM-5 Development., 2010), Future research should attempt to examine the latent structure of MDD in relation to the structure of other mood and anxiety disorders, such as dysthymic disorder, PTSD, and generalized anxiety disorder, given their conceptual and empirical overlap with MDD (Watson, 2005, 2009). Although recent studies have examined the underlying factors accounting for diagnostic variables marking these disorders (Forbes et al., 2010, 2011), future work should attempt to analyze this issue with these disorders at the latent construct level (e.g., Elhai et al., 2011).

Several limitations apply to the present study. First, we used the PHQ-9 which is a self-report instrument, albeit administered via telephone format in our study. We do not know the extent to which our findings with a self-report instrument would replicate using a comprehensive, structured diagnostic interview to query MDD, such as the Structured Clinical Interview for *DSM-IV*'s (SCID) depression modules, given the difference in instrument format. Second, our study sampled Army National Guard military personnel (predominantly male), rather than a broader, representative community or clinical sample of men and women. Future work should explore MDD's factor structure using more generalizable civilian and military samples. Third, our sample as a whole reported only mild depression in their lifetimes, and thus we were unable to examine depression's factor structure among a heterogeneous sample varying on current depression severity. Finally, future research should test relationships between MDD factors and specific external constructs of psychopathology.

Author note

This project was funded by the Department of Defense Congressionally Directed Medical Research Program W81XWH-07-1-0409, the "Combat Mental Health Initiative."

Correspondence about this article may be addressed to Dr. Elhai through his website: 〈www.jon-elhai.com〉

Disclosures

The authors have no perceived or actual conflicts of interest related to this project. The sponsor had no role in study design, data collection, analysis, interpretation of results, report writing or manuscript submission.

References

American Psychiatric Association. DSM-5 Development., 2010. Proposed Draft Revisions to DSM Disorders and Criteria.

Baas, K.D., Cramer, A.O.J., Koeter, M.W.J., van de Lisdonk, E.H., van Weert, H.C., Schene, A.H., 2011. Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). Journal of Affective Disorders 129, 229–235.

Bollen, K.A., 1989. Structural Equations with Latent Variables. John Wiley & Sons.

Cameron, I.M., Crawford, J.R., Lawton, K., Reid, I.C., 2008. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. British Journal of General Practice 58, 32–36.

Cannon, D.S., Tiffany, S.T., Coon, H., Scholand, M.B., McMahon, W.M., Leppert, M.F., 2007. The PHQ-9 as a brief assessment of lifetime major depression. Psychological Assessment 19, 247–251.

Dum, M., Pickren, J., Sobell, L.C., Sobell, M.B., 2008. Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addictive Behaviors 33, 381–387.

Elhai, J.D., Contractor, A.A., Palmieri, P.A., Forbes, D., Richardson, J.D., 2011. Exploring the relationship between underlying dimensions of posttraumatic stress disorder and depression in a national, trauma-exposed military sample. Journal of Affective Disorders 133, 477–480.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J., 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 4, 272–399.

Fan, X., Sivo, S.A., 2009. Using Δgoodness-of-fit indexes in assessing mean structure invariance. Structural Equation Modeling 16, 54–67.

Flynn, H.A., Sexton, M., Ratliff, S., Porter, K., Zivin, K., 2011. Comparative performance of the Edinburgh Postnatal Depression Scale and the Patient Health Questionnaire-9 in pregnant and postpartum women seeking psychiatric services. Psychiatry Research 15, 130–134.

Forbes, D., Lockwood, E., Elhai, J.D., Creamer, M., O'Donnell, M., Bryant, R., McFarlane, A., Silove, D., 2011. An examination of the structure of posttraumatic stress disorder in relation to the anxiety and depressive disorders. Journal of Affective Disorders 132, 165–172.

Forbes, D., Parslow, R., Creamer, M., O'Donnell, M., Bryant, R., McFarlane, A., Silove, D., Shalev, A., 2010. A longitudinal analysis of posttraumatic stress disorder

symptoms and their relationship with fear and anxious–misery disorders: implications for DSM-V. Journal of Affective Disorders 127, 147–152.
Grieger, T.A., Cozza, S.J., Ursano, R.J., Hoge, C., Martinez, P.E., Engel, C.C., Wain, H.J., 2006. Posttraumatic stress disorder and depression in battle-injured soldiers. American Journal of Psychiatry 163, 1777–1783.
Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling 6, 1–55.
Huang, F.Y., Chung, H., Kroenke, K., Delucchi, K.L., Spitzer, R.L., 2006. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. Journal of General Internal Medicine 21, 547–552.
Kalpakjian, C.Z., Toussaint, L.L., Albright, K.J., Bombardier, C.H., Krause, J.K., Tate, D.G., 2009. Patient Health Questionnaire-9 in spinal cord injury: an examination of factor structure as related to gender. The Journal of Spinal Cord Medicine 32, 147–156.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Kline, R.B., 2010. Principles and Practice of Structural Equation Modeling, 3rd edn. Guilford Press.
Krause, J.S., Bombardier, C., Carter, R.E., 2008. Assessment of depressive symptoms during inpatient rehabilitation for spinal cord injury: is there an underlying somatic factor when using the PHQ? Rehabilitation Psychology 53, 513–520.
Krause, J.S., Reed, K.S., McArdle, J.J., 2010. Factor structure and predictive validity of somatic and nonsomatic symptoms from the Patient Health Questionnaire-9: a longitudinal study after spinal cord injury. Archives of Physical Medicine and Rehabilitation 91, 1218–1224.
Kroenke, K., Spitzer, R.L., Williams, J.B., Lowe, B., 2010. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. General Hospital Psychiatry 32, 345–359.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine 16, 606–613.
Muthén, B.O., Muthén, L., 2006. Chi-square Difference Testing Using the Satorra–Bentler Scaled Chi-square. Muthén & Muthén.
Olsen, L.R., Jensen, D.V., Noerholm, V., Martiny, K., Bech, P., 2003. The internal and external validity of the major depression inventory in measuring severity of depressive states. Psychological Medicine 33, 351–356.
Raftery, A.E., 1995. Bayesian model selection in social research. Sociological Methodology 25, 111–163.
Richardson, E.J., Richards, J.S., 2008. Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. Rehabilitation Psychology 53, 243–249.
Satorra, A., Bentler, P.M., 2001. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66, 507–514.
Shafer, A.B., 2006. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. Journal of Clinical Psychology 62, 123–146.
Spitzer, R.L., Williams, J.B.W., Kroenke, K., Linzer, M., Ill, d.F.V., Hahn, S.R., Brody, D., Johnson, J.G., 1994. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. Journal of American Medical Association 272, 1749–1756.
Wang, P.S., Demler, O., Olfson, M., Pincus, H.A., Wells, K.B., Kessler, R.C., 2006. Changing profiles of service sectors used for mental health care in the United States. American Journal of Psychiatry 163, 1187–1198.
Watson, D., 2005. Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. Journal of Abnormal Psychology 114, 522–536.
Watson, D., 2009. Differentiating the mood and anxiety disorders: a quadripartite model. Annual Review of Clinical Psychology 5, 221–247.