Journal of Affective Disorders 165 (2014) 166–169

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Brief report

Comparing four competing models of depressive symptomatology: A confirmatory factor analytic study of 986,647 U.S. veterans

Jack Tsai [a,b,*], Jon D. Elhai [c], Robert H. Pietrzak [b,d], Rani A. Hoff [b,e], Ilan Harpaz-Rotem [b,d]

[a] United States Department of Veterans Affairs New England Mental Illness Research, Education, and Clinical Center, 950 Campbell Avenue, 151D, West Haven, CT 06516, USA
[b] Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06510, USA
[c] Departments of Psychology and Psychiatry, University of Toledo, Toledo, OH 63606, USA
[d] United States Department of Veterans Affairs National Center for Posttraumatic Stress Disorder, Clinical Neurosciences Division, VA Connecticut Healthcare System, West Haven, CT 06516, USA
[e] Veterans Affairs Northeast Program Evaluation Center, West Haven, CT 06516, USA

ARTICLE INFO

Article history:
Received 20 April 2014
Accepted 25 April 2014
Available online 4 May 2014

Keywords:
Depression
Veterans
Confirmatory factor analyses

ABSTRACT

Background: Few rigorous studies have examined the factor structure of major depression symptoms as assessed by current diagnostic systems. This study evaluated four competing models of depressive symptomatology among a large, heterogeneous sample of U.S. veterans.
Methods: To determine the best fitting model of major depressive symptoms among four competing models, this study conducted a series of confirmatory factor analyses on a national sample of 986,647 U.S. veterans.
Results: A two-factor model first reported by Krause, Reed, and McArdle (2010) provided superior fit to symptom-level data compared to three other models. The optimal model consists of a somatic factor including anhedonia, sleep difficulties, fatigue, appetite changes, concentration difficulties, and psychomotor agitation; and a non-somatic factor including depressed mood, feelings of worthlessness, and thoughts of death. Factorial invariance testing found this model to be invariant by gender and major depression diagnosis.
Limitations: A widely used self-report measure of depression was used and the sample consisted solely of veterans so further study is needed with clinician-administered measures and non-veteran samples.
Conclusions: Together, these findings support separating symptoms of major depression into somatic and non-somatic factors which may have clinical relevance, and help clarify debates about the factor structure of depressive symptoms.

Published by Elsevier B.V.

1. Introduction

Major depressive disorder (MDD) is among the most prevalent and commonly diagnosed psychiatric disorders (Grant et al., 2009; Hasin et al., 2005; Kessler et al., 2003, 1994). However, little is known about the symptom structure of MDD as assessed by diagnostic criteria in the most recent version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) and its predecessor, the DSM-IV (American Psychiatric Association, 2000). Understanding the structure of MDD symptoms is important because it may inform approaches to the assessment, treatment, and neurobiology of this disorder (Antonijevic, 2008; Hasler et al., 2004; Henningsen et al., 2003; Sullivan et al., 2000).

A widely used screening instrument for MDD that maps directly onto DSM-IV and DSM-5 criteria is the Patient Health-Questionnaire-9 (PHQ-9; Kroenke et al., 2001). There is currently debate about the factor structure of MDD as assessed by the PHQ-9. Four competing, empirically-supported models have emerged (Table 1). The first labeled Model 1 is a one-factor model, in which all MDD symptoms represent one latent construct. Three two-factor models (Model 2a–c) have been found, all containing latent constructs that can be labeled somatic (e.g., sleep difficulties, appetite changes, and fatigue) and non-somatic (e.g., depressed mood, feelings of worthlessness, and thoughts of death) factors. However, these three two-factor models differ with respect to the factor assignment of three MDD symptoms, namely, anhedonia, concentration difficulties, and psychomotor agitation/retardation.

* Corresponding author at: United States Department of Veterans Affairs New England Mental Illness Research, Education, and Clinical Center, 950 Campbell Avenue, 151D, West Haven, CT 06516, USA. Tel.: +1 203 932 5711x2090.
E-mail address: Jack.Tsai@yale.edu (J. Tsai).

http://dx.doi.org/10.1016/j.jad.2014.04.075
0165-0327/Published by Elsevier B.V.

Table 1
Four competing factor models of the Patient Health Questionnaire-9 (PHQ-9) items.

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. Feelings 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 1 proposes one latent factor and Models 2a–2c propose two latent factors. Model 2a was originally reported by Krause et al. (2008). Model 2b was originally reported by Richardson and Richards (2008) and 2c was reported by Krause et al. (2010).

Concentration difficulties and psychomotor agitation/retardation are assigned to a non-somatic factor in Model 2a (Krause et al., 2008), but to a somatic factor in Models 2b (Richardson and Richards, 2008) and 2c (Krause et al., 2010). Anhedonia is assigned to a non-somatic factor in Models 2a and 2b, but to a somatic factor in Model 2c. Thus, there lacks empirical consensus on the optimal model of MDD symptoms. Knowing which symptoms cluster with which somatic or non-somatic symptoms may contribute to knowledge about depressive symptomatology and its etiology, and may have implications for treatments targeted at certain symptom clusters.

A recent study of 2615 soldiers in the Ohio Army National Guard compared these four models using confirmatory factor analysis (CFA) and found that Model 2c provided the best fit to PHQ-9 data (Elhai et al., 2012). However, this finding has not been replicated on a large national heterogeneous sample and the factorial invariance of Model 2c with respect to gender and diagnostic status has not been evaluated. CFA is one of the best statistical tools to compare competing structural models as, unlike exploratory factor analysis, it provides fit indices based on pre-specified and theory-driven models.

Using CFAs, we evaluated four competing models of MDD symptoms using the PHQ-9 in a sample of nearly one million U.S. veterans. We then examined factorial invariance of the optimal model with respect to gender and clinician-assessed MDD diagnosis. We hypothesized that Model 2c would provide a superior fit compared to the other models and that this model would demonstrate factorial invariance by gender and MDD diagnosis.

2. Methods

2.1. Sample

Using U.S. Department of Veterans Affairs (VA) electronic medical record databases that capture outpatient care and test results, we identified all veterans who received mental health care and completed a PHQ-9 between October 2007 and September 2012. The data were unduplicated to include only the first record. The data were unduplicated to include only the first record so that there were no repeated measurements for any individual veteran. The study sample included 986,647 unique veterans from 130 VA facilities across the country.

Demographic information and diagnoses of major depression for each veteran were also obtained from the electronic medical record. Diagnoses of major depression were made and documented by VA clinicians in the medical record within a year of the initial administration of the PHQ-9. Of the study sample, 91.71% were male, 60.67% were White, 51.46% were married, 12.76% served in Operations Iraqi Freedom/Enduring Freedom, and 8.08% had a clinician-assessed MDD diagnosis.

2.2. Assessment

The Patient Health-Questionnaire-9 (PHQ-9; Kroenke et al., 2001) is a 9-item multipurpose instrument for screening, diagnosing, monitoring, and measuring the severity of depression. The PHQ-9 incorporates DSM-IV and DSM-5 diagnostic criteria for MDD. Respondents are asked to rate the frequency to which they experience various symptoms, such as “little interest or pleasure in doing things,” “poor appetite or overeating”, or “thoughts that you would be better off dead or of hurting yourself in some way.” Response options range from 0 (not at all) to 3 (nearly every day). The PHQ-9 has been found to have strong psychometric properties, including diagnostic validity and reliability (Kroenke et al., 2001).

2.3. Data analysis

First, we conducted CFAs to examine absolute model fit of the four models. Model fit was evaluated using the Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Akaike Information Criterion (AIC), Sample size-adjusted Bayesian Information Criterion (BIC), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) values. Model fit was determined using empirically-defined benchmarks, as follows: CFI and TLI $\geq .95$ indicative of excellent fit; RMSEA $\leq .08$ for adequate model fit; and SRMR $\leq .05$ for excellent fit (Hu and Bentler, 1999); and lower AIC and sample-size adjusted BIC values.

Second, we compared the four competing models using robust maximum likelihood estimation with the Satorra–Bentler chi-square (S–B χ^2) scaling correction, robust to non-normality (Satorra and Bentler, 2001). PHQ-9 items were specified to load on only one factor, factors were allowed to correlate, all error covariances were fixed to zero, and all tests were two-tailed. S–B χ^2 difference tests for nested models were used to compare the relative fit of the one-factor model to the three two-factor models (Fan and Sivo, 2009). The different two-factor models, which are non-nested, were compared on fit indices. A 10-point BIC difference between models represents a 150:1 likelihood and “very strong” ($p < .05$) support that the model with the smaller BIC value fits best (Kass and Raftery, 1995).

Third, we tested the factorial invariance of the optimal model. The total sample was divided into three random samples and CFAs were repeated on each random sample. CFAs were also repeated on subsamples of only female veterans and only veterans with a diagnosis of clinician-assessed MDD diagnoses. Factorial invariance tests were conducted for gender and clinician-assessed MDD diagnosis at the configural, metric, and scalar levels. The configural level requires the same items to load on the same factors across groups. The metric level further requires equivalence of factor loadings for items across groups. The scalar level requires both

Table 2
Fit indices from confirmatory factor analyses of different models using the Patient Health Questionnaire-9 (n= 986,647).

Model	S–B χ^2	df	CFI	TLI	RMSEA	AIC	Adjusted-BIC	SRMR
Model 1	149,337.45	27	.96	.94	.08	21,761,426.73	21,761,659.45	.03
Model 2a	117,690.39	26	.96	.95	.07	21,706,370.11	21,706,611.45	.03
Model 2b	141,108.30	26	.96	.94	.07	21,746,425.06	21,746,666.41	.03

Note: Model 1 proposes one latent factor and Models 2a–2c propose two latent factors. Model 2a was originally reported by Krause et al. (2008). Model 2b was originally reported by Richardson and Richards (2008) and 2c was reported by Krause et al. (2010). S–B χ2= Satorra–Bentler chi-square; df= degrees of freedom; CFI= Comparative Fit Index; TLI= Tucker Lewis Index; RMSEA= Root Mean Square Error of Approximation; AIC= Akaike Information Criterion; Adjusted-BIC= Sample size-adjusted Bayesian Information Criterion; SRMR= Standardized Root Mean Square Residual.

equivalence of factor loadings and item intercepts across groups. Factorial invariance testing often relies on chi-square difference testing, which is affected by sample size (our study uses an unusually large sample size) and unequal sample sizes can bias results (Brown, 2006). Thus, it is recommended that difference in CFI values between models be evaluated as this fit index has been found to be superior to chi-square difference testing; a change in CFI of .01 or less is indicative of invariance (Cheung and Rensvold, 2002). All analyses were conducted with Mplus version 7.1 (Muthén and Muthén, 2012).

3. Results

As shown in Table 2, all four competing models showed good-to-excellent fit to PHQ-9 symptom-level data, as evidenced by standard benchmarks (CFI and TLI $\geq$.95 and RMSEA $\leq$.08, and SRMR $\leq$.05). S–B χ^2 difference tests revealed that compared to the Model 1, Model 2a had a significantly better fit, $\Delta X^2(1)$= 1383121.13, $p<.001$; as did Model 2b, $\Delta X^2(1)$= 8229.15, $p<.001$; and Model 2c, $\Delta X^2(1)$= 31647.06, $p<.001$. Therefore, all three two-factor models showed a significantly better fit than the one-factor model.

Comparing the two-factor models, Model 2c had better fit indices, as indicated by higher TLI values and lower S–B χ^2, AIC, and BIC values relative to Models 2a and 2b. Moreover, there was a greater than 10,000-point BIC difference between Model 2c and Models 2a and 2b, providing "very strong" support favoring Model 2c (a 10-point BIC difference represents a 150:1 likelihood).

When the total sample was divided in three random subsamples, CFAs repeated on each of the three subsamples yielded similar results favoring Model 2c (results available upon request from author). When CFAs were conducted on only subsamples of female veterans (n=81,327) and veterans with a MDD diagnosis (n=79,263), the results again favored Model 2c. Further, tests of factorial invariance for both gender and MDD diagnosis at the configural, metric, and scalar levels found CFI values of.97,.96, and.96, respectively; this change in CFI values of .01 or less demonstrates invariance at the scalar level.

4. Discussion

Using a national sample of nearly one million U.S. veterans, we found that DSM-5 MDD symptoms are best represented by two factors, a somatic and non-somatic (or affective) factor. Furthermore, we compared three competing two-factor models and found the model referred to as Model 2c (Krause et al., 2010) provided superior fit to PHQ-9 data. In this model, the somatic factor includes sleep difficulties, fatigue, appetite changes, concentration difficulties, anhedonia, and psychomotor agitation/retardation. The non-somatic or affective factor includes depressed mood, feelings of worthlessness, and thoughts of death. This finding replicates and extends the results of a previous CFA study on a smaller, geographically limited sample (Elhai et al., 2012).

Furthermore, we demonstrated that this factor structure is invariant by gender and clinician-assessed MDD diagnoses. Model 2c remained the optimal model when the study sample was constrained to only female veterans and to only veterans with MDD. These findings contribute to our understanding of the nature of DSM-5 MDD symptoms and help clarify debates about which of the four competing models best represents depressive symptomatology.

Clinicians treating depression should consider the somatic and non-somatic components of the disorder, and the symptoms that underlie these components. For example, concentration difficulties may reflect restlessness more than affect, and thus may require a different treatment approach. Different symptom expressions may also affect treatment engagement and outcomes. Patients with more somatic complaints may be more likely to be seen in primary care compared to those who have non-somatic or affective complaints (Kirmayer et al., 1993; Simms et al., 2012). But further research is needed on the clinical implications of conceptualizing MDD as consisting of somatic and non-somatic factors. To further validate Model 2c, additional studies are needed to examine how the somatic and non-somatic factors of Model 2c are differentially related to treatment engagement, functional outcomes, neurobiological substrates, and the phenotype of MDD.

4.1. Limitations

This study had a few limitations worth noting. The PHQ-9 is a self-report measure and we do not know whether there is a method variance if a clinician-administered instrument was used instead. This study also exclusively sampled veterans, as did a previous study that compared these four competing models (Elhai et al., 2012), so generalizability to the general adult population is unknown. There were several strengths of the study including the use of a large sample size with varying depression symptom severity; repeated CFA results on randomly divided samples and subsamples; factorial invariance testing; and the extension of previous findings.

Role of funding source
No funding source had a role in this study.

Conflict of interest
Disclosures: None of the authors report any conflicts of interest related to this work. Dr. Tsai has received funding from the Bristol Meyers Squibb Foundation, which had no influence on this work. This work was supported by the United States Department of Veterans Affairs, Office of Research and Development. The views presented here are those of the authors alone and do not represent the position of any federal agency or of the United States Government.

References

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. American Psychiatric Association, Washington, DC.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Publishing, Arlington, VA.

Antonijevic, I., 2008. HPA axis and sleep: identifying subtypes of major depression. Stress: Int. J. Biol. Stress 11, 15–27.

Brown, T.A., 2006. Confirmatory Factor Analysis for Applied Research. Guilford Press, New York.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model. 9, 233–255.

Elhai, J.D., Contractor, A.A., Tamburrino, M., Fine, T.H., Prescott, M.R., Shirley, E., Chan, P.K., Slembarski, R., Liberzon, I., Galea, S., 2012. The factor structure of major depression symptoms: a test of four competing models using the Patient Health Questionnaire-9. Psychiatry Res. 199, 169–173.

Fan, X., Sivo, S.A., 2009. Using goodness-of-fit indexes in assessing mean structure invariance. Struct. Equ. Model. 16, 54–67.

Grant, B.F., Goldstein, R.B., Chou, S.P., Huang, B., Stinson, F.S., Dawson, D.A., Saha, T. D., Smith, S.M., Pulay, A.J., Pickering, R.P., 2009. Sociodemographic and psychopathologic predictors of first incidence of DSM-IV substance use, mood and anxiety disorders: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. Mol. Psychiatry 14, 1051–1066.

Hasin, D.S., Goodwin, R.D., Stinson, F.S., Grant, B.F., 2005. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. Arch. Gen. Psychiatry 62, 1097–1106.

Hasler, G., Drevets, W.C., Manji, H.K., Charney, D.S., 2004. Discovering endophenotypes for major depression. Neuropsychopharmacology 29, 1765–1781.

Henningsen, P., Zimmermann, T., Sattel, H., 2003. Medically unexplained physical symptoms, anxiety, and depression: a meta-analytic review. Psychosom. Med. 65, 528–533.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). J. Am. Med. Assoc. 289, 3095–3105.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C., Hughes, M., Eschleman, S., Wittchen, H.U., Kendler, K.S., 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the US: results from the National Comorbidity Survey. Arch. Gen. Psychiatry 51, 8–19.

Kirmayer, L.J., Robbins, J.M., Dworkind, M., Yaffe, M.J., 1993. Somatization and the recognition of depression and anxiety in primary care. Am. J. Psychiatry 150, 734–741.

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? Rehabil. Psychol. 53, 513.

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. Arch. Phys. Med. Rehabil. 91, 1218–1224.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.

Muthén, L.K., Muthén, B.O., 2012. Mplus User's Guide. Muthén & Muthén, Los Angeles, CA.

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. Rehabil. Psychol. 53, 243.

Satorra, A., Bentler, P.M., 2001. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66, 507–514.

Simms, L.J., Prisciandaro, J.J., Krueger, R.F., Goldberg, D.P., 2012. The structure of depression, anxiety and somatic symptoms in primary care. Psychol. Med. 42, 15.

Sullivan, P.F., Neale, M.C., Kendler, K.S., 2000. Genetic epidemiology of major depression: review and meta-analysis. Am. J. Psychiatry 157, 1552–1562.