Changes in the Beck Depression Inventory-II's Underlying Symptom Structure Over 1 Month of Inpatient Treatment

Jon D. Elhai, PhD,† Ateka A. Contractor, MA,* Tracey L. Biehn, MA,* Jon G. Allen, PhD,‡§ John Oldham, MD,‡§ Julian D. Ford, PhD,|| Anouk L. Grubaugh, PhD,¶ and B. Christopher Frueh, PhD‡#*

Abstract: Research has not investigated changes in the symptom structure of depression over the course of mental health treatment. In the present study, 1025 psychiatric inpatients were recruited and assessed for depression symptom severity using the Beck Depression Inventory-II (BDI-II) at admission and after 1 month of treatment. A three-factor BDI-II model was tested using confirmatory factor analysis and fit reasonably well at both time points. Measurement invariance testing results demonstrated that factor loadings increased, indicating that the meaning of the three underlying depression dimensions changed through treatment. However, observed variable intercepts and residual error variances decreased significantly after 1 month of treatment, reflecting decreases in symptom severity as well as measurement error. Thus, depressive symptom severity decreased over the course of treatment, and the underlying factor structure of depression improved in fit after treatment. Implications for changes to the structure of depression symptoms and in the clinical practice of tracking depression over time are discussed.

Key Words: Depression, factor analysis, Beck Depression Inventory, inpatients, psychiatric treatment

(*J Nerv Ment Dis* 2013;201: 371–376)

Several widely used depression symptom instruments have been analyzed to identify the latent dimensions of depression. Much of this previous work has been conducted using the Center for Epidemiological Studies–Depression Scale (CES-D) (*e.g.*, Elhai et al., 2011; Nguyen et al., 2004; Shafer, 2006; Sheehan et al., 1995; Van Dam and Earleywine, 2011; Wood et al., 2010) and Patient Health Questionnaire-9 (*e.g.*, Baas et al., 2011; Kalpakjian et al., 2009; Krause et al., 2010; Richardson and Richards, 2008). Recent studies have examined depression's underlying structure using the Beck Depression Inventory-II (BDI-II), finding support for several two- and three-factor depression models (discussed below). However, investigators have not examined changes in depression's symptom structure through the course of mental health treatment.

In the present study, we examined changes in the BDI-II's factor structure over a 1-month course of inpatient mental health treatment. The original BDI was developed in the 1960s to quantify the severity of depression with a self-report format (Beck et al., 1961). It was subsequently revised in 1979 (BDI-IA) and then again in 1996 (BDI-II). The BDI-II has 21 items (Beck et al., 1996b) and is more consistent with the item content of *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition's* (*DSM-IV*) major depressive episode symptoms than was the original BDI. Specifically, four new items of agitation, concentration difficulties, worthlessness, and loss of energy replaced the original BDI's items of somatic preoccupation, weight loss, body image concerns, and difficulty at work. Furthermore, items tapping sleep and appetite impairment were modified to include either hyperexperience or hypoexperience of those symptoms (Beck et al., 1996b).

Factor analytic research of the BDI-II has found little support for a one-factor model of depression (Dum et al., 2008; Siegert et al., 2009). Rather, most studies support some variation of a two-factor model involving somatic and cognitive factors (Arnau et al., 2001; Dozois et al., 1998; Steer et al., 2000; Vanheule et al., 2008; Whisman et al., 2000) or a three-factor model of somatic, cognitive, and affective factors (Buckley et al., 2001; Johnson et al., 2006; Osman et al., 1997; Seignourel et al., 2008; Steer et al., 1998), and these models were tested with a diverse set of samples, including medical patients, college students, outpatient mental health, and substance use patients. However, it should be noted that in a minority of studies, support has been found for higher order models, including an intercorrelated model with a second-order depression factor (Grothe et al., 2005; Steer et al., 1999) and a bifactor model defined by lower order specific factors and a lower order general depression factor (Ward, 2006). Thus, factor analytic research suggests that the BDI-II is a multidimensional instrument of depression; however, some debate exists concerning its precise underlying factors.

The purpose of the current study was to assess the extent to which depression's factor structure, as measured by the BDI-II, is invariant over a 1-month course of inpatient mental health treatment. This issue is important in understanding how basic psychopathological processes of depression may change over time from treatment. Furthermore, the decision to select a 1-month measurement point was based on two statistical reasons. First, there was a significant drop-off in subjects completing assessments after the 1-month measurement point, reducing the sample size and power to detect significant results if subsequent time points were used. Second, given that there was a difference in the total BDI-II scores from baseline to the 1-month measurement point (detailed below), it adds to the validity of assessing the symptom structure difference between the two time points.

Given the diversity of depression factor analytic models mentioned above, the present study considered only models supported by multiple studies. Three depression models met this criterion, including the three-factor model of cognitive, affective, and somatic dimensions of Buckley et al. (2001) (further supported by Johnson et al., 2006; Seignourel et al., 2008; Vanheule et al., 2008), Ward's (2006) bifactor model of cognitive and somatic dimensions (supported by Quilty et al., 2010), and the three-factor model of Osman et al. (1997) discussed below (supported by Quilty et al., 2010) (see Table 1 for these models' item mappings). We decided not to use the model

*Departments of Psychology and †Psychiatry, University of Toledo, Ohio; ‡The Menninger Clinic; §Department of Psychiatry, Baylor College of Medicine, Houston, TX; ||Department of Psychiatry, University of Connecticut Health Center, Farmington, CT; ¶Department of Psychiatry, Medical University of South Carolina, Charleston, SC; and #Department of Psychology, University of Hawaii, Hilo.

Correspondence about this article may be addressed to Jon Elhai through his Web site: www.jon-elhai.com.

Send reprint requests to Jon D. Elhai, PhD, Department of Psychology, University of Toledo, Mail Stop 948, 2801 W Bancroft St, Toledo, OH 43606-3390.

ISSN: 0022-3018/13/20105–0371
DOI: 10.1097/NMD.0b013e31828e1004

TABLE 1. Replicated Models of the BDI-II Factor Structure

BDI-II Item	Buckley et al. (2001)	Osman et al. (1997)	Ward (2006)
Sadness	Cognitive	Negative attitude	–
Pessimism	Cognitive	Negative attitude	Cognitive
Past failure	Cognitive	Negative attitude	Cognitive
Loss of pleasure	Affective	Performance difficulty	–
Guilty feelings	Cognitive	Negative attitude	Cognitive
Punishment feelings	Cognitive	Negative attitude	Cognitive
Self-dislike	Cognitive	Negative attitude	Cognitive
Self-criticalness	Cognitive	Negative attitude	Cognitive
Suicidal thoughts	Cognitive	Negative attitude	Cognitive
Crying	Affective	Negative attitude/somatic	–
Agitation	Somatic	Somatic	–
Loss of interest	Affective	Performance difficulty	–
Indecisiveness	Affective	Performance difficulty	–
Worthlessness	Cognitive	Negative attitude	Cognitive
Loss of energy	Somatic	Performance difficulty	Somatic
Changes in sleeping patterns	Somatic	Somatic	Somatic
Irritability	Somatic	Performance difficulty	–
Changes in appetite	Somatic	Somatic	Somatic
Concentration difficulty	Somatic	Performance difficulty	Somatic
Tiredness or fatigue	Somatic	Performance difficulty	Somatic
Loss of interest in sex	Somatic	Somatic	–

In Ward's model, all items additionally loaded on a G factor (general) (Ward, 2006).
BDI-II indicates Beck Depression Inventory-II.

of Buckley et al. because studies supporting its use relied solely on restricted and nonrelevant samples of substance abusers, which were not the target sample in this current study. Furthermore, we omitted analysis of the bifactor model of Ward et al. because the complex nature of its hierarchical model made it impossible to statistically analyze measurement invariance across time.[1] As a result, in the present study, we tested for invariance of the BDI-II using the empirically supported three-factor model of depression of Osman et al., over a 1-month course of mental health treatment.

The three-factor oblique model of Osman et al. (1997) includes the following intercorrelated depression factors: Negative Attitude, Performance Difficulty, and Somatic problems. The Negative Attitude factor includes several affective components such as guilty feelings, sense of failure, and punishment feelings. The Performance Difficulty factor taps such items as loss of interest, concentration problems, and loss of energy. Finally, the Somatic Elements factor taps somatic components of depression. In this model, the crying item is specified to cross-load onto both the Negative Affect and Somatic Elements factors (Osman et al., 1997); furthermore, for three pairs of items, residual error variances are specified to covary.

In addition, the focus on the Osman model has statistical and methodological reasons. First, studies establishing and validating Osman's model have used the statistically superior confirmatory factor analysis (CFA) (Osman et al., 1997; Quilty et al., 2010) compared with the arguably inferior exploratory factor analysis (EFA), adding to the confidence and validity of the results. To elaborate, CFA does not have the same limitations of EFA; CFA is not exploratory and thus does not capitalize on chance error while assessing patterns of item correlations in the data but rather tests a priori hypothesized models of interest and often generalizes to other samples (Fabrigar et al., 1999). Second, the Osman model was supported in studies using samples similar to the one used in the current study. To elaborate, the sample used in the Quilty et al. (2010) study was one of the few samples using adults with clinically significant depressive symptoms compared with a general primary care sample or substance-related sample (Buckley et al., 2001; Johnson et al., 2006).

We sampled psychiatric inpatients using the BDI-II and then implemented invariance testing procedures to examine the extent to which depression's factor structure is stable through a 1-month course of treatment. Specifically, we examined differences over time in factor loadings (indicating the extent to which depression factors have the same meaning over time), intercepts (estimating item severity), and residual variance unaccounted for by the factors. Sheehan et al. (1995) conducted a relevant study, examining changes over time in the factor structure of the CES-D among a sample of patients with rheumatoid arthritis. The authors tested for changes in factor loadings and residual error variances (but not intercepts), revealing that these parameters were invariant over time. In the present study, we hypothesized that the factor structure would be invariant across time in psychiatric treatment, except for differences expected in observed variable intercepts as a result of reduced symptom severity from treatment. This study is important in understanding how the structure of depression symptoms may be stable or change over the course of treatment.

METHODS

Participants

Participants were psychiatric patients enrolled at The Menninger Clinic, a private, not-for-profit facility that offers intensive inpatient mental health treatment in Houston, TX, first founded in 1919. Participants were consecutively enrolled in treatment between April 2008 and December 2010. For a given patient who had multiple admissions during the period, data from his/her admission(s) subsequent to the first were excluded. Treatment is diverse and comprehensive at the clinic and includes evidence-based interventions that

[1]We attempted to test this model using measurement invariance procedures but encountered estimation problems resulting from empirical underidentification.

broadly incorporate medication management, addiction services (when applicable), tailored individual and group psychotherapy and psychoeducation, and structured interpersonal and recreational activities. Individual therapy involved mainly cognitive-behavioral and psychodynamic orientations and was provided twice weekly by psychologists (but occasionally by psychiatrists and social workers). Furthermore, process-oriented group psychotherapy addressed patients' interactions and relationships in the group. Coupled with psychotherapy, patients met regularly with their psychiatrist and core team members in clinical rounds, participated in family meetings with their social worker, attended multiple nursing assessments and ad hoc meetings with members of the nursing staff daily, and attended a range of psychoeducational and activities groups throughout the week.

A total of 1488 participants were sampled. However, 463 of these subjects were discharged before the 1-month BDI-II follow-up, leaving an effective sample size of 1025 participants treated for at least 1 month. Among the effective sample, 527 (51.4%) were women and 498 (48.6%) were men. Age ranged from 18 to 78 years, averaging 34.73 years (SD, 14.14 years). Most participants reported being unemployed (n = 618, 60.4%) or working at least 30 hours per week (n = 231, 22.6%), and the majority reported being never married (n = 585, 57.6%) or currently married (n = 274, 27.0%). Representative of most private inpatient mental health facilities, most participants were Caucasian (n = 930, 92.1%). Thirty-six subjects (3.6%) identified as being of Hispanic or Latino ethnicity. A significant number of participants reported some college education (n = 122, 35.8%), a bachelor's degree (n = 96, 28.2%), or a master's degree (n = 35, 10.3%). Of these demographic variables, discharged patients and the effective sample's patients differed only on age (with dropouts being older) ($F_{1,\ 1486} = 14.73$, $p < 0.001$; $\eta_p^2 = 0.01$), although the effect was small.

In addition, the treating psychiatrist diagnosed patients at the time of admission on the basis of the history of the present illness, a psychiatric history of the patient and family, a psychosocial history, and a mental status examination. The psychiatric diagnoses were updated throughout treatment and final diagnoses are made at the time of discharge, informed by updated information from the patient and family throughout treatment as well as clinical observations made by the multidisciplinary staff during the course of treatment. Based on the aforementioned unstandardized diagnostic interviews, using a recent clinic archival records search, the most prevalent category of diagnosis found (albeit with missing data for 293 subjects) was alcohol use disorder (n = 417, 34.9%), bipolar disorder (n = 157, 13.1%), and psychotic disorder (n = 101, 8.5%). However, limitations of unstructured interviews must be kept in mind when assessing the usefulness of the aforementioned diagnostic information, mainly less comprehensiveness, less validity, more bias (Garb, 2007), lower interrater reliabilities (Miller, 2001b), and lower diagnostic accuracy than structured interviews (Miller, 2001a).

The archival chart review did not include information on major depressive disorder diagnostic prevalence. However, our emphasis is on the range of depression severity rather than clinically significant depression because our focus is on individual differences in expression of depression severity. This focus is also consistent with research demonstrating that depression is more of a continuous variable rather than a categorical one (Hankin et al., 2005).

Procedure

Data were collected as part of the clinic's Adult Outcomes Project, described in detail elsewhere (Allen et al., 2009). In brief, all participants were assessed using established, validated measures at admission and periodically over the course of treatment. Assessment was conducted via a hospital-wide Web survey on laptop computers. This project was a clinical outcomes project, conducted with all patients; thus, no patients declined participation, as it was part of their routine clinical care. Use of the project's data was approved by Baylor College of Medicine's Institutional Review Board. Several measures were administered to patients during the course of their treatment (Allen et al., 2009). Of relevance to the present study, we discuss the assessment of depression using the BDI-II below.

Beck Depression Inventory-II

The BDI-II (Beck et al., 1996b) is a 21-item self-report measure of depression symptoms, modified from the original BDI to be more consistent with *DSM-IV* major depressive disorder item content. The BDI-II demonstrates adequate test-retest reliability (r = 0.93), internal consistency (alphas of 0.91–0.93 with clinical and nonclinical samples), and convergent/discriminant construct validity with external measures of depression and anxiety (Beck et al., 1996a, 1996b; Dozois et al., 1998). In the current study, internal consistency (alpha) ranged from 0.92 (baseline) to 0.93 (follow-up).

Analysis

There were 1025 subjects who were still hospitalized at the 1-month time point whose data were used to examine the 1-month invariance of depression's factor structure. There were no missing data values because of the "forced response" Web survey design used. Univariate nonnormality was found among the BDI-II items, with one item evidencing skewness of 2.23 and kurtosis of 5.87. Multivariate nonnormality was also revealed (Mardia's skewness $z = 21.49$, $p < 0.001$, and Mardia's kurtosis $z = 540.61$, $p < 0.001$).

In addition, several other statistical assumptions precede the use of CFA. In terms of multicollinearity of the data, wherein the item correlations within a factor are assessed, this assumption is violated if the correlations are 0.80 or higher (Tabachnick and Fidell, 2007). Thus, our results showing the highest correlations of 0.7 indicate that the aforementioned assumption is met. Furthermore, CFA requires sufficient sample size for adequate power, and the current study sample size meets the recommended guidelines (Muthén and Muthén, 2002). Lastly, the model is mathematically identified, meaning that it is theoretically possible to assess the hypothesized pattern of correlations and covariances in the data (Kline, 2010).

We specified our model based on the model of Osman et al. (1997), including that model's specification of a cross-factor-loading for the crying item. However, in the interest of parsimony and preserving low type I error, we did not implement the residual error covariance estimation of Osman et al. between several items; arbitrary (nontheoretically meaningful) covarying of residual error variances is noted as a substantial limitation in factor analytic research (reviewed in Kline, 2010). Thus, all residual covariances were fixed to zero. CFAs used maximum likelihood estimation, treating the BDI items as continuously scaled, thus analyzing a Pearson's covariance matrix. We used a mean-adjusted Satorra-Bentler (S-B) chi-square statistic, which is robust to nonnormality (Satorra and Bentler, 2001). Goodness-of-fit indices are reported, 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 that fit very well (or adequately) are indicated by CFI and TLI greater than 0.95 (0.90–0.94), RMSEA less than 0.06 (to 0.08), and SRMR less than 0.08 (to 0.10) (Hu and Bentler, 1999). Differences between models in Bayesian Information Criterion (BIC) values of more than 10 points favor the model with the smaller BIC value (Kass and Raftery, 1995).

We first assessed model fit based on the model of Osman et al. (1997) for the baseline and 1-month BDI-II administrations separately. We subsequently conducted measurement invariance testing using parameter equality constraints with this model, testing invariance/noninvariance across the baseline and 1-month item sets on factor loadings, observed variable intercepts, and observed variable measurement errors, following established procedures (*e.g.*, Meredith and Teresi, 2006). Invariance testing was conducted to test potential differences in parameter estimates between the time points.

In the invariance testing, model A included all 42 items (21 baseline BDI-II items and 21 1-month BDI-II items), specified to load on their respective factors from the Osman et al. (1997) model (detailed in the *Introduction*). Thus, we allowed a particular time point's specific item response parameter (*e.g.*, baseline item 1's factor loading) to freely vary from that of the other time point's item response parameter (*e.g.*, 1-month follow-up item 1's factor loading). Subsequent models tested progressively more conservative restrictions, constraining particular parameter estimates to be equal across time points, tested against the previous step's model. Model B constrained factor loadings as equal across time points (testing metric or pattern invariance). Model C additionally constrained observed variable intercepts to be equal (testing strong or scalar factorial invariance). Last, model D additionally constrained residual variances as equal (testing strict factorial invariance).

Tests of statistical significance between models were assessed with chi-square difference tests, comparing a given CFA model assuming equal parameter estimates (*e.g.*, factor loadings) against a model allowing those estimates to vary across time points. A correction factor was used for taking the difference between nonnormally distributed S-B chi-square values to approximate normally distributed chi-square values (Muthén and Muthén, 2006; Satorra and Bentler, 2001). Because invariance testing involved factor loadings, we estimated all loadings in our models (rather than using the traditional method of fixing a factor's unstandardized loading to 1), assigning factor metrics instead by fixing their variances to 1. All analyses were conducted using Mplus 6 software.

RESULTS

Total scores on the BDI-II ranged from 0 to 59 for the baseline administration (mean [SD], 26.10 [12.86]) and from 0 to 58 for the 1-month follow-up (mean [SD], 16.37 [12.06]). As expected, a significant difference was found, indicating that depression scores significantly decreased after a month of treatment ($F_{1,\,1024} = 811.66$, $p < 0.001$, partial $\eta^2 = 0.44$).

CFA for the baseline administration of the BDI-II (using the Osman model) resulted in S-B χ^2_{185} ($N = 1025$) = 927.32, $p < 0.001$, CFI = 0.91, TLI = 0.90, RMSEA = 0.06, SRMR = 0.04, BIC = 51227.99. In contrast, CFA for the follow-up administration of the BDI-II (using the Osman model) resulted in S-B χ^2_{185} ($N = 1025$) = 742.41, $p < 0.001$, CFI = 0.93, TLI = 0.93, RMSEA = 0.05, SRMR = 0.04, BIC = 45527.26. Thus, based on differences in fit indices (as outlined above), the 1-month administration seemed to fit slightly better.

Measurement invariance analyses, using the Osman et al. (1997) model, formally tested differences between baseline and 1-month follow-up BDI-II item sets on the structural parameters. Based on difference tests (see Table 2: "All Factors" column), analyses revealed that factor loadings were significantly different between the baseline and 1-month administrations (model A versus B); this was determined by a significantly worse fit when constraining loadings to be equal between the baseline and 1-month administrations. Specifically, factor loadings were larger on average for the 1-month administration. In addition, intercepts (model B versus C) and residual variances (model C versus D) were significantly different across administrations. In particular, intercepts and residual variances were higher, on average, at the baseline administration.

On the basis of the results obtained above, we conducted additional analyses to determine which of the BDI-II's factors were responsible for the differences between administrations. We reconducted measurement invariance analyses, each implementing equality constraints across the administrations (*i.e.*, for loadings, intercepts, residual variances); we implemented three sets of these analyses separately—one set for each of the BDI-II's three factors, testing invariance for only one factor at a time. Table 2 indicates that each factor was significantly different between the administrations.

DISCUSSION

In the present study, we analyzed differences in the factor structure of the BDI-II between baseline and 1 month after inpatient mental health treatment. We found that factor loadings increased, whereas intercepts and residual error variances decreased significantly over the 1-month course of treatment. Thus, overall, it seems that the BDI-II's factor structure was not invariant across the 1-month period of treatment, which previous clinic data show was a period of significant improvement for most patients (Latini et al., 2009).

Obtaining smaller item intercepts after 1 month of treatment would be expected, as this indicates that depression was endorsed at lower levels after this 1-month time point. This finding is consistent with research demonstrating a reduction in depressive symptoms after inpatient mental health treatment (*e.g.*, Hennings et al., 2009; Schneider et al., 2005), in contrast to findings showing limited evidence of improvement in the full spectrum of treatments for depression (Pettit et al., 2009). Noteworthy, however, was that factor loadings increased over the course of treatment. Thus, it seems that after a reduction in depression symptoms over 1 month of treatment, the BDI-II's latent factors became better defined and captured more variance in the observed items. This finding is supported by the fact that residual error variances decreased over time, and model fit was slightly better at the 1-month time point.

The present study's findings demonstrate that not only does the BDI-II's depression severity decrease with treatment but also that the instrument's factor structure changes. The change does not seem to involve a shift in the relationship among depression symptoms as reported by inpatients, such as a clearer distinction between mood-related and more physiological symptoms, as was found for depressed versus

TABLE 2. Comparisons of Baseline and 1-Month Follow-up Equality Constraints Using Measurement Invariance Procedures

Chi-square Difference Test				
Models Tested	**All Factors Constrained**	**Negative Attitudes Constrained**	**Performance Constrained**	**Somatic Constrained**
A versus B	χ^2_{Diff} = 81.20 (22)**	χ^2_{Diff} = 42.88 (10)**	χ^2_{Diff} = 26.17 (7)**	χ^2_{Diff} = 11.19 (5)*
B versus C	χ^2_{Diff} = 4689.49 (21)**	χ^2_{Diff} = 2014.56 (10)**	χ^2_{Diff} = 2273.53 (7)**	χ^2_{Diff} = 523.56 (5)**
C versus D	χ^2_{Diff} = 456.23 (21)**	χ^2_{Diff} = 249.85 (10)**	χ^2_{Diff} = 139.72 (7)**	χ^2_{Diff} = 81.00 (5)**

The table presents corrected versions of the difference between Satorra-Bentler chi-square values for the model comparisons; degrees of freedom are in parentheses. Model A = no parameters constrained to be equal across administrations; model B = factor loadings constrained to be equal; model C = observed variable intercepts and factor loadings constrained to be equal; model D = residual variances, factor loadings, and observed variable intercepts constrained to be equal.

*$p < 0.05$.

**$p < 0.001$.

nondepressed adults by Santor and Coyne (2001). Instead, after 1 month of treatment, the BDI-II symptom items tend to load more strongly on each factor, suggesting that depression symptoms may become more coherently interrelated within domains (*i.e.*, cognitive, behavioral, somatic) and more distinct between domains after depression and other emotional distress are somewhat more under control. That is, high levels of emotional distress that precipitate the hospital admission may create extraneous variance or noise in these patients' symptom ratings that artifactually decrease how well depression's factor structure fits the data.

The findings suggest that to obtain precise depression testing results from which to infer patients' clinical status, clinicians may need to treat test results taken immediately upon admission as more indicative of acute distress and instability than as evidence of the underlying severity of psychopathology or of prognosis. Further research could target assessing the changes in BDI-II symptom ratings over a longer period, with an aim to assess the function of this questionnaire for tracking diagnostic severity and treatment progress. This may be particularly important to do at follow-ups after discharge from inpatient treatment, to determine if the structure of depression changes or remains invariant when the added stressors and reduced supports of life outside a sheltered inpatient treatment milieu are encountered. Future research could also examine invariance of the depression factor structure among individuals who do not receive treatment to determine to what extent changes observed in the BDI-II factor structure in this study are caused by treatment rather than recovery associated with the natural passage of time.

Several limitations apply to the present study. First, our findings are generalizable only to the BDI-II and not to other self-report or interview-based measures of depression. Second, we analyzed only those patients who remained in treatment over the course of 1 month, and it is unclear if and how including patients who left treatment before this time point would have altered the current study findings.

CONCLUSIONS

This study provides the first empirical evidence that the structure of depression symptoms may change over time as a result of treatment while also demonstrating that this change is caused by an increased coherence of the contribution of symptoms to factors rather than to a fundamental shift in the tripartite structure of depressive symptoms. Study findings point toward further study of how intensive treatment not only reduces the initial severity of acute depressive symptoms but also may help to clarify the nature of those symptoms apart from the generalized emotional distress that occurs when depression is acutely severe.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

Allen JG, Frueh BC, Ellis TE, Latini DM, Mahoney JS, Oldham JM, Sharp C, Wallin L (2009) Integrating outcomes assessment and research into clinical care in inpatient adult psychiatric treatment. *Bull Menninger Clin*. 73:259–295.

Arnau RC, Meagher MW, Norris MP, Bramson R (2001) Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychol*. 20:112–119.

Baas KD, Cramer AOJ, Koeter MWJ, van de Lisdonk EH, van Weert HC, Schene AH (2011) Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). *J Affect Disord*. 129:229–235.

Beck AT, Steer RA, Ball R, Ranieri WF (1996a) Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *J Pers Assess*. 67:588–597.

Beck AT, Steer RA, Brown GK (1996b) *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. *Arch Gen Psychiatry.* 4:561–571.

Buckley TC, Parker JD, Heggie J (2001) A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. *J Subst Abuse Treat*. 20:197–204.

Dozois DJA, Dobson KS, Ahnberg JL (1998) A psychometric evaluation of the Beck Depression Inventory-II. *Psychol Assess*. 10:83–89.

Dum M, Pickren J, Sobell LC, Sobell MB (2008) Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. *Addict Behav.* 33:381–387.

Elhai JD, Contractor AA, Palmieri PA, Forbes D, Richardson JD (2011) Exploring the relationship between underlying dimensions of posttraumatic stress disorder and depression in a national, trauma-exposed military sample. *J Affect Disord*. 133:477–480.

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*. 4: 272–399.

Garb HN (2007) Computer-administered interviews and rating scales. *Psychol Assess*. 19:4–13.

Grothe KB, Dutton GR, Jones GN, Bodenlos J, Ancona M, Brantley PJ (2005) Validation of the Beck Depression Inventory-II in a low-income African American sample of medical outpatients. *Psychol Assess*. 17:110–114.

Hankin BL, Fraley RC, Lahey BB, Waldman ID (2005) Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *J Abnorm Psychol*. 114:96–110.

Hennings JM, Owashi T, Binder EB, Horstmann S, Menke A, Kloiber S, Dose T, Wollweber B, Spieler D, Messer T, Lutz R, Knzel H, Bierner T, Pollmcher T, Pfister H, Nickel T, Sonntag A, Uhr M, Ising M, Holsboer F, Lucae S (2009) Clinical characteristics and treatment outcome in a representative sample of depressed inpatients: Findings from the Munich Antidepressant Response Signature (MARS) project. *J Psychiatr Res*. 43:215–229.

Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equation Model*. 6:1–55.

Johnson ME, Neal DB, Brems C, Fisher DG (2006) Depression as measured by the Beck Depression Inventory-II among injecting drug users. *Assessment*. 13:168–177.

Kalpakjian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG (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 RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc*. 90:773–795.

Kline RB (2010) *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Krause JS, Reed KS, McArdle JJ (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.

Latini DM, Allen JG, Seo M, Mahoney JS, Ellis TE, Frueh BC (2009) Psychosocial characteristics of psychiatric inpatients at admissions and discharge: The Menninger Clinic Adult Outcomes Project. *Bull Menninger Clin*. 73:296–310.

Meredith W, Teresi JA (2006) An essay on measurement and factorial invariance. *Med Care*. 44:S69–S77.

Miller PR (2001a) Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. *Psychiatry Res*. 105:255–264.

Miller PR (2001b) Inpatient diagnostic assessments: 2. Interrater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Res*. 105:265–271.

Muthén BO, Muthén L (2006) *Chi-square difference testing using the Satorra-Bentler scaled chi-square* (Vol. 2007). Los Angeles. CA: Muthén & Muthén.

Muthén LK, Muthén BO (2002) How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equation Model*. 4:599–620.

Nguyen HA, Kitner-Triolo M, Evans MK, Zonderman AB (2004) Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Res*. 126:177–187.

Osman A, Downs WR, Barrios FX, Kopper BA, Gutierrez PM, Chiros CE (1997) Factor structure and psychometric characteristics of the Beck Depression Inventory-II. *J Psychopathol Behav Assess*. 19:359–376.

Pettit JW, Lewinsohn PM, Roberts RE, Seeley JR, Monteith L (2009) The long-term course of depression: Development of an empirical index and identification of early adult outcomes. *Psychol Med*. 39:403–412.

Quilty LC, Zhang KA, Bagby RM (2010) The latent symptom structure of the Beck Depression Inventory-II in outpatients with major depression. *Psychol Assess*. 22:603–608.

Richardson EJ, Richards JS (2008) Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. *Rehabil Psychol*. 53:243–249.

Santor DA, Coyne JC (2001) Evaluating the continuity of symptomatology between depressed and nondepressed individuals. *J Abnorm Psychol*. 110:216–225.

Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 66:507–514.

Schneider F, Hrter M, Brand S, Sitta P, Menke R, Hammer-Filipiak U, Kudling R, Heindl A, Herold K, Frommberger U, Elmer O, Hetzel G, Witt G, Wolfersdorf M, Berger M, Gaebel W (2005) Adherence to guidelines for treatment of depression in in-patients. *Br J Psychiatry.* 187:462–469.

Seignourel PJ, Green C, Schmitz JM (2008) Factor structure and diagnostic efficiency of the BDI-II in treatment-seeking substance users. *Drug Alcohol Depend*. 93:271–278.

Shafer AB (2006) Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol*. 62:123–146.

Sheehan TJ, Fifield J, Reisine S, Tennen H (1995) The measurement structure of the Center for Epidemiologic Studies Depression Scale. *J Pers Assess*. 64:507–521.

Siegert RJ, Walkey FH, Turner-Stokes L (2009) An examination of the factor structure of the Beck Depression Inventory-II in a neurorehabilitation inpatient sample. *J Int Neuropsychol Soc*. 15:152–157.

Steer RA, Ball R, Ranieri WF, Beck AT (1999) Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *J Clin Psychol*. 55:117–128.

Steer RA, Kumar G, Ranieri WF, Beck AT (1998) Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *J Psychopathol Behav Assess*. 20:127–137.

Steer RA, Rissmillera DJ, Beck AT (2000) Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behav Res Ther*. 38:311–318.

Tabachnick BG, Fidell LS (2007) *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.

Van Dam NT, Earleywine M (2011) Validation of the Center for Epidemiologic Studies Depression Scale—Revised (CESD-R): Pragmatic depression assessment in the general population. *Psychiatry Res*. 186:128–132.

Vanheule S, Desmet M, Groenvynck H, Rosseel Y, Fontaine J (2008) The factor structure of the Beck Depression Inventory-II: An evaluation. *Assessment*. 15:177–187.

Ward LC (2006) Comparison of factor structure models for the Beck Depression Inventory-II. *Psychol Assess*. 18:81–88.

Whisman MA, Perez JE, Ramel W (2000) Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. *J Clin Psychol*. 58:545–551.

Wood AM, Taylor PJ, Joseph S (2010) Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. *177:120–123.*