Psychological Assessment
2014, Vol. 26, No. 4, 1106–1115

1040-3590/14/$12.00 http://dx.doi.org/10.1037/a0036998

Factor Structure and Diagnostic Validity of the Beck Depression Inventory–II With Adult Clinical Inpatients: Comparison to a Gold-Standard Diagnostic Interview

Andrew M. Subica
Loyola Marymount University

J. Christopher Fowler
The Menninger Clinic, Houston, Texas

Jon D. Elhai
University of Ohio, Toledo

B. Christopher Frueh
Baylor College of Medicine and University of Hawaii at Hilo

Carla Sharp
University of Houston

Erin L. Kelly
University of Southern California

Jon G. Allen
The Menninger Clinic, Houston, Texas

Little is known about the psychometric properties and clinical utility of the Beck Depression Inventory–II (BDI-II) among adult clinical inpatients, a group at high risk for major depressive disorder (MDD). Data from 1,904 adult inpatients were analyzed using confirmatory factor analysis (CFA), Cronbach's alpha, and Pearson's correlations. Receiver operating characteristic (ROC) analyses evaluating MDD diagnostic performance were conducted with a subsample ($n = 467$) using a structured diagnostic interview for reference. CFA of 3 previous 2-factor oblique solutions, observed in adolescent and older adult inpatient clinical samples, and 3 corresponding bifactor solutions indicated that BDI-II common item variance was overwhelmingly accounted for by 1 general factor specified to all items, with minor additional variance contributed by 2 specific factors. Analyses revealed high internal consistency (Cronbach's $\alpha = .93$) and significant ($p < .01$) intercorrelations between the BDI-II total scale and Behavior and Symptom Identification Scale–24's Depression/Functioning ($r = .79$) and Overall ($r = .82$) subscales. ROC analyses generated low area under the curve (.695; 95% confidence interval [.637, .752]) and cutoff scores with poor sensitivity/specificity balance. BDI-II use as a screening instrument for overall depressive symptomology was supported, but MDD diagnostic performance was suboptimal. Clinicians are advised to use the BDI-II to gauge severity of depression and measure clinical changes to depressive symptomology over time but to be mindful of the limitations of the BDI-II as a diagnostic tool for adult inpatients.

Keywords: major depressive disorder, psychiatric hospital, psychometrics, reliability, ROC curve

Major depressive disorder (MDD) is a pervasive, often disabling mental disorder that is common (16.2% lifetime prevalence rate; Kessler et al., 2003); highly comorbid with other mental disorders (Currie et al., 2005; Kessler et al., 1996); and significantly impairs mental, physical, and psychosocial function (Johnson, Weissman, & Klerman, 1992; Wells et al., 1989). MDD has been linked to elevated psychiatric hospitalization rates (Costello, 1982) due to the presence of depressed mood, cognitive difficulties, somatic symptoms, mood instability, and suicidal thoughts (American Psychiatric Association, 2000). In clinical populations, MDD is associated with elevated suicide mortality (Angst, Stassen, Clayton, & Angst, 2002; Simon & VonKorff, 1998), triggering the need to

This article was published Online First June 16, 2014.

Andrew M. Subica, Psychology Applied Research Center, Loyola Marymount University; J. Christopher Fowler, The Menninger Clinic, Houston, Texas; Jon D. Elhai, Department of Psychology, University of Ohio, Toledo; B. Christopher Frueh, Baylor College of Medicine, and Department of Psychology, University of Hawaii at Hilo; Carla Sharp, Department of Psychology, University of Houston; Erin L. Kelly, School of Social Work, University of Southern California; Jon G. Allen, The Menninger Clinic, Houston, Texas.

Andrew M. Subica's contribution to this project was partially supported by Grant UL1RRO31986 from the National Center for Research Resources of the National Institutes of Health. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health.

Correspondence concerning this article should be addressed to Andrew M. Subica, 1 LMU Drive, University Hall, Suite 4726, Los Angeles, CA 90045. E-mail: subica@gmail.com

develop efficient and reliable means of assessing MDD in these individuals. Self-report depression screening tools such as the widely used Beck Depression Inventory, (2nd ed.; BDI-II; Beck, Steer, & Brown, 1996) are intended to assist clinicians in determining which adults admitted for a psychiatric crisis require a comprehensive depression assessment and provide insight into the severity of patients' depression (Weissman et al., 1977). Unfortunately, it is presently unclear whether the BDI-II is appropriate for screening depressive symptomology and MDD in adult clinical inpatients (despite its substantial potential for improving the recognition and subsequent treatment of MDD for this high-need population) due to a lack of empirical research.

The purpose of this study was to conduct the first large-scale examination of the MDD symptom and diagnostic screening capabilities of the BDI-II in adults receiving inpatient psychiatric care. The BDI-II, an update to the well-validated Beck Depression Inventory (BDI-IA; Beck, Steer, Ball, & Ranieri, 1996; Beck, Steer, & Carbin, 1988; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was revised to match the major depressive episode diagnostic criteria in the *Diagnostic and Statistical Manual for Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 2000). Major changes included (1) adding four new symptoms (i.e., worthlessness, difficulty concentrating, loss of energy, and agitation), (2) removing four BDI-IA symptoms (i.e., change in body image, somatic preoccupation, weight loss, and work difficulty), and (3) extending the symptom timeframe from 1 to 2 weeks.

Previous BDI-II Factorial Findings

Since the BDI-II's release, its psychometric characteristics have been well documented across diverse community and clinical populations including primary care patients (Arnau, Meagher, Norris, & Bramson, 2001), adult clinical outpatients (Beck, Steer, et al., 1996; Steer, Ball, Ranieri, & Beck, 1997), adolescent clinical inpatients (Dolle et al., 2012; Krefetz et al., 2002; Osman, Kopper, Barrios, Gutierrez, & Bagge, 2004), and older adult clinical inpatients (Steer, Rissmiller, & Beck, 2000), although not adult clinical inpatients. Studies examining the BDI-II's factorial structure have regularly reported a robust two-factor oblique model (Beck, Steer, & Brown, 1996; Dozois et al., 1998; Steer et al., 2000; Storch, Roberti, & Roth, 2004; Whisman, Perez, & Ramel, 2000) composed of cognitive and noncognitive/affective symptoms, with between-studies variation in the specific item composition of the two correlated factors. Less frequently, alternative factor structures, including one-factor (Steer, Clark, Beck, & Ranieri, 1999) and three-factor (Buckley, Parker, & Heggie, 2001; Lindsay & Skene, 2007) models, have also been reported.

The factorial data consistently affirm the BDI-II as multidimensional, assessing multiple domains of depressive symptomology with heterogeneous indicators. Yet, the recommended clinical scoring of the BDI-II is unidimensional, by calculating the patient's total scale score as an index of severity of general depression rather than domain-specific subscales (Beck, Steer, & Brown, 1996; Quilty et al., 2010; Reise, Moore, & Haviland, 2010; Ward, 2006). Reasons for this discrepancy are multifactorial but are principally explained by two research findings. First, multidimensional BDI-II factors tend to be highly correlated with poor discrimination of scale items across factors (Ward, 2006). For example, the cognitive factor overlaps with a prominent self-critical depression subtype that is clearly associated with depressive affect (Blatt, 2004). Therefore, partitioning the BDI-II into cognitive and noncognitive/affective subscales would present limited clinical value due to the substantial overlap between the latent factors. Second, BDI-II common item variance appears to be independently accounted for by both a general depression factor and multiple specific factors (Brouwer et al., 2013; Quilty, Zhang, & Bagby, 2010; Ward, 2006), raising a critical question: For adult clinical inpatients, how does one determine the appropriate factor structure *and* clinical interpretability of an instrument that appears to be comprised of items that simultaneously assess a general depression construct and narrower subdomains of depressive symptomology?

One solution to address this dual methodological and clinical question is bifactor modeling, a contemporary factor analytic approach to evaluating psychological instruments for which common item variation appears to reflect variation on both a single construct (i.e., unidimensional total scale score) and multiple constructs (i.e., multidimensional subscales; Chen, West, & Sousa, 2006; Reise et al., 2010; Reise, Morizot, & Hays, 2007). To briefly review relevant BDI-II factor analytic strategies, determining the constructs that underlie common item variation involves testing theoretically and/or data-driven unidimensional, correlated traits, second-order, or bifactor structural models. A unidimensional BDI-II model parsimoniously tests whether one common source of variance exists for all 21 items but fails to account for additional systematic sources of variance for each item (Reise et al., 2010). Thus, multidimensional "correlated traits" BDI-II models are tested to evaluate whether BDI-II common item variance is better described by multiple correlated specific factors ("traits"). Problematically, these models fail to account for the general depression factor, missing a potentially critical source of item variance. Second-order BDI-II models address this problem by testing whether correlated specific factors share a common underlying higher order factor, that is, a general depression factor. This general depression factor is conceptualized as explaining the variation between correlated specific factors rather than variation between scale items, and for this reason, second-order models are ideal when higher order factors are hypothesized to account for the relationship among lower order specific factors (Chen et al., 2006).

In contrast, bifactor models examine the common variation between BDI-II items independently explained by orthogonal general depression and specific factors and are therefore optimal for separately testing the unidimensionality (general depression factor) and multidimensionality (specific factors) of the BDI-II. By calculating the contributions of specific factors to common item variance above and beyond the general factor, a challenge for second-order models, bifactor models help to demonstrate the utility of subscale scores independent of the total score. Consequently, prior researchers have opted to test BDI-II data using bifactor instead of second-order factor models and have found consistently improved model fit for bifactor versus typical correlated traits models (Brouwer et al., 2013; Osman, Barrios, Gutierrez, Williams, & Bailey, 2008; Quilty et al., 2010; Vanheule, Desmet, Groenvynck, Rosseel, & Fontaine, 2008; Ward, 2006). Further description of the relative advantages of bifactor versus second-order factor models for psychological measures can be found in Chen et al. (2006).

Presently, we know of only three prior studies that have applied bifactor modeling to BDI-II clinical data, with none analyzing data from adult inpatients. The first study (Ward, 2006) used two previously tested oblique two-factor models and a third bifactor model to explore the BDI-II's structure in three clinical outpatient and two college samples. The two-factor oblique models were drawn from Beck, Steer, and Brown's (1996) original two-factor cognitive and somatic–affective model extracted from 500 clinical outpatients and from Whisman, Perez, and Ramel's (2000) two-factor cognitive–affective and somatic model confirmed with 576 college students. The bifactor model specified orthogonal general depression, somatic, and cognitive factors and contained two correlated errors. Ward (2006) discovered that the bifactor model provided equivalent or better fit to item data for all five samples and that the general factor primarily accounted for common item variance, diminishing the additive clinical value of the subscales.

Quilty et al. (2010) examined the fit of six two-factor and three three-factor oblique models, and Ward's (2006) bifactor orthogonal model, to data from 425 clinical outpatients with MDD. The bifactor model produced the best model fit, whereas only one of the oblique models adequately fit sample data. Based on the best fit of Ward's bifactor model combined with observations of strong correlations between specific factors in the oblique models, Quilty et al. concluded that their results justified using the BDI-II's total scale score but not its subscales.

Most recently, Brouwer et al. (2013) compared the fit of a unidimensional model, two two-factor oblique models and their corresponding bifactor models, and Ward's (2006) bifactor model in a sample of 1,530 clinical outpatients. As expected, the fit of all three bifactor models was superior to that of the unidimensional and two-factor oblique models, with the general factor accounting for 74%–77% of the common item variance (compared to 3%–16% for the specific factors) across all bifactor models, strongly supporting the clinical use of the total scale score as opposed to subscale scores. Based on the collective evidence in favor of bifactor modeling for clinical BDI-II data, we chose to test bifactor models using our adult clinical inpatient data.

Factor Structures of the BDI-II Among Clinical Inpatients

To our knowledge, no studies of adult inpatients have examined bifactor BDI-II models, and only one prior study has sought to validate the BDI-II with this population (J. C. Cole, Grossman, Prilliman, & Hunsaker, 2003). J. C. Cole et al.'s (2003) validation study assessed 101 adult inpatients, reporting strong internal consistency (Cronbach's alpha = .95), moderate correlation with the Grossman–Cole Depression Inventory (r = .73; J. C. Cole et al., 2003), and support for a two-factor cognitive and noncognitive model via maximum-likelihood estimation (MLE) extraction with direct oblimin rotation. Unfortunately, the precise factor structure evaluated via confirmatory factor analysis was not reported by the authors and could not be tested.

Therefore, to frame our current analyses, we used two studies that reported BDI-II factor structures for adolescent or older adult clinical inpatients. One study of 408 adolescent inpatients (Osman et al., 2004) reported poor fit of previously published one-, two-, and three-factor models including Beck, Steer, and Brown's (1996) original two-factor oblique model to their data, leading them to extract from sample data a two-factor oblique model containing cognitive–affective and somatic factors via MLE extraction with promax rotation. For older adult inpatients, Steer et al.'s (2000) study of 130 geriatric (ages 55 and older) inpatients with depression found high internal consistency (Cronbach's alpha = .90) and evidence for a two-factor cognitive and noncognitive model using MLE extraction with promax rotation. We replicated analyses of these two oblique factor models (and compared them with their bifactor counterparts) in the current study, as we theorized these models, extracted from alternative inpatient samples, would more closely fit our adult inpatient sample than would earlier models extracted from clinical outpatient or community samples (due to their presumably lower psychopathology). For comparative purposes, we also examined the well-established original two-factor oblique BDI-II model (Beck, Steer, & Brown, 1996) and its corresponding bifactor model.

The present investigation sought to address the lacuna in BDI-II literature of adult clinical inpatients. Our first research objective was to determine the optimal BDI-II factor structure for adult clinical inpatients by comparing existing correlated traits models observed in alternative clinical populations and their corresponding bifactor models using a large sample of adult inpatients. Our second objective was to evaluate the internal consistency, and convergent and discriminant characteristics of BDI-II scale scores conforming to our best fitting factor models. The final objective was to examine the BDI-II's clinical utility as an MDD diagnostic screen among adult inpatients based on prior evidence indicating the BDI-II's diagnostic effectiveness for screening MDD in adolescent inpatients (Dolle et al., 2012; Krefetz et al. 2002).

Method

Sampling and Participants

The study sample (see Allen et al., 2009, for initial methods description) was composed of 1,904 adults (M = 34.78 years, SD = 14.53; 50.4% women) admitted to the Menninger Clinic for intensive psychiatric treatment from April 2008 to May 2012. BDI-II findings from a subset of the current sample were previously reported in a study that examined changes in the symptom structure of depression during a 1-month course of treatment using measurement invariance testing (Elhai et al., 2013), as opposed to the current investigation of the psychometric and diagnostic properties of the BDI-II; 910 participants (47.79%) overlapped between the two samples. Participants' mean length of stay was 41.02 days (SD = 21.31). Ethnocultural composition of the sample was 81.36% (n = 1,549) White; 4.31% (n = 82) multiracial; 1.41% (n = 27) Asian; 0.84% (n = 16) Native Hawaiian, Pacific Islander, American Indian, or Alaskan Native; 0.42% (n = 8) African American; and 11.66% (n = 222) missing or preferring not to answer. When queried about Hispanic versus non-Hispanic heritage, 65 participants (3.41%) further reported being of Hispanic heritage. Within the sample, 24.89% (n = 474) of the participants were married or living with someone as married; 63.71% (n = 1,213) were single, separated, divorced, or widowed; and 11.40% (n = 217) had missing data or preferred not to answer. With respect to prior mental health treatment, participants reported seeing an average of 3.76 (SD = 4.62) therapists and 2.91 (SD = 3.17) psychiatric medication providers prior to the current admis-

sion. Participants reported a mean of 2.65 (SD = 8.91) previous admissions for acute/crisis (1–5 days) and extended (>5 days) psychiatric hospital care.

A subsample of 575 participants (M = 33.74 years, SD = 14.07; 53.7% women) consisting of consecutive patients admitted between October 2010 and May 2012 completed the Structured Clinical Interview for *DSM* Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 2002), to replace a less-structured clinical interview, in accordance with new hospital diagnostic protocols. The mean length of stay for participants in this subsample was 47.69 days (SD = 17.36). The ethnocultural composition of the subsample was 89.56% (n = 515) White; 5.91% (n = 34) multiracial; 1.22% (n = 7) Asian; 0.70% (n = 4) Native Hawaiian, Pacific Islander, American Indian, or Alaskan Native; 0.87% (n = 5) African American; and 1.74% (n = 10) missing or preferring not to answer. In response to an item specifically addressing Hispanic heritage, 28 (4.87%) participants indicated Hispanic heritage. In terms of marital status, 24.52% (n = 141) were married or living as married; 73.91% (n = 425) were single, separated, divorced, or widowed; and 1.57% (n = 9) had missing data or preferred not to answer. Participants reported an average of 2.49 (SD = 7.25) previous admissions for acute/crisis and extended psychiatric hospital care.

Participants completed two self-report study measures—the BDI-II and the Behavior and Symptom Identification Scale-24 (BASIS-24; Eisen, Normand, Belanger, Spiro, & Esch, 2004)—upon admission. An additional 575 participants were also diagnosed by trained assessors at admission using the SCID-I. Although all participants completed the BDI-II and BASIS-24 at 2-week treatment intervals and at discharge, only admission data are reported. Study protocols were approved by the institutional review board of the Baylor School of Medicine.

Measures

BDI-II. The BDI-II is a 21-item self-report measure of depression that asks respondents to rate the severity of their depressive symptoms during the past 2 weeks using a variable Likert scale (i.e., 19 items use a 4-point scale, two items use a 7-point scale). Individual item scores are summed to create a total severity score with a range of 0 to 63. Total scores can be used to categorize respondents by depressive severity using the following ranges: 0 to 13 (minimal); 14 to 19 (mild); 20 to 28 (moderate); >28 (severe; Beck, Steer, & Brown, 1996). Cutoff scores for presumptive diagnosis of MDD of 23 (Dolle et al., 2012) and 24 (Krefetz et al., 2002) have been identified for adolescent but not adult clinical inpatients.

BASIS-24. The BASIS-24 (Eisen et al., 2004) is a validated 24-item self-report measure intended to cut across psychiatric diagnoses to assess diverse psychopathology during treatment (Cameron et al., 2007). An abbreviated version of the BASIS-32 (Eisen, Wilcox, Leff, Schaefer, & Culhane, 1999), the BASIS-24 features seven subscales: Overall, Depression/Functioning, Relationships, Self-Harm, Emotional Lability, Substance Abuse, and Psychosis. Items are rated on a 5-point Likert scale, with higher scores indicating greater symptom severity. Scale scores are calculated by multiplying scores for each item by its weight and summing the weighted ratings for all questions in the measure (for the Overall score) or all questions in each subscale (for subscale scores) using a scoring algorithm. Factor analyses of the BASIS-24 with multiple clinical populations have confirmed the six subscales (Eisen, Gerena, Ranganathan, Esch, & Idiculla, 2006), but at present, the Overall summary score has only been confirmed using item-response theory analyses (Eisen et al., 2004). The Overall score measures general mental health distress (Eisen et al., 1999), and the Depression/Functioning subscale measures depression and anxiety symptomology, as well as daily role functioning. We judged these two subscales to be conceptually similar to potential BDI-II constructs and, therefore, analyzed them in this study to determine the relatedness of BDI-II scores and these subscales. The Substance Abuse (examining possible drug and alcohol problems) and Psychosis (examining cardinal psychotic symptoms) subscales, on the other hand, we judged to be least conceptually aligned with potential BDI-II constructs and explored their distinctiveness from possible BDI-II scores. The remaining subscales were not analyzed.

SCID-I. The SCID-I is a semistructured diagnostic interview validated for use in research and clinical settings (First et al., 2002). The interview assesses Axis I disorders listed in the *DSM–IV* and contains symptom criteria and specifiers for disorder subtype, course, and severity. In the present study, the SCID-I was used to identify MDD prevalence in a subsample (n = 575).

Statistical Analysis

Confirmatory factor analyses (CFA) were conducted using IBM SPSS Amos Version 21 (Arbuckle, 2012). Descriptive statistics, reliability analyses, t tests, and correlations were conducted with the IBM Statistical Package for Social Sciences (SPSS) Version 21. Receiver operating characteristic (ROC) analyses were evaluated via SPSS Version 21 and confirmed using MedCalc 12.7.5 (MedCalc, 2013).

CFA of prior psychiatric patient models. The plausibility of all factor solutions to the sample covariance matrix (N = 1,904) were evaluated using maximum-likelihood CFA with 10,000 Bollen–Stine bootstrap samples applied for inference to accommodate our multivariate nonnormally distributed data. Maximum-likelihood estimation has been shown to be robust to deviations from normal distribution for larger sample sizes (Browne & Shapiro, 1988; Finch, West, & MacKinnon, 1997; Satorra & Bentler, 1990), especially when accompanied by Bollen–Stine bootstrap, a modified procedure superior to naïve bootstrapping for calculating model test statistics and fit indexes (Bollen & Stine, 1992). No correlated errors were permitted within tested factor solutions, and all factor variances were constrained to one.

To establish model fit, the chi-square goodness-of-fit test was calculated but not used, given its sensitivity to minor discrepancies between proposed and null models, leading to false rejection of the null hypothesis when analyzing large samples (D. A. Cole, 1987; Floyd & Widaman, 1995). Instead, goodness of fit was determined using the following fit indices: comparative fit index (CFI), standardized root-mean-square residual (SRMR), and root-mean-square error of approximation (RMSEA). Conventional cutoff criteria indicating acceptable model fit for these indexes are ≥.90 for CFI, ≤.08 for SRMR, and ≤.08 for RMSEA (Browne & Cudeck, 1993; Kline, 2005; Schreiber, Nora, Stage, Barlow, & King, 2006); ≥.95 for CFI and ≤.07 for RMSEA signify good model fit (Hu & Bentler, 1999; Steiger, 2007). Within the bifactor

solutions, the independent contributions of general and specific factors to common item variance were determined by calculating the percentage of explained common variance (ECV) for each factor. This statistic was derived by summing the squared standardized regression weights (factor loadings) for each factor divided by the sum of all squared standardized regression weights in the total solution (Brouwer et al., 2013).

Unidimensional. A solution with all 21 items loaded onto one factor.

Beck, Steer, and Brown (1996). The original two-factor oblique solution described in the BDI-II manual (Beck, Steer, and Brown, 1996) for adult outpatients was tested. Model factors were cognitive (Items 1–3, 5–9, 14) and somatic–affective (Items 4, 10–13, 15–21).

Osman et al. (2004). A two-factor oblique solution consisting of cognitive–affective (Items 1–9, 12–14) and somatic (Items 11, 15–21) factors observed in adolescent inpatients was fit to study data. In this solution, Item 10 (Crying) failed to load on either factor.

Steer et al. (2000). A two-factor oblique solution observed in geriatric inpatients composed of cognitive (Items 1, 2, 4, 11–12, 15, 17–21) and noncognitive (3, 5–10, 13, 14) factors was tested against study data. Item 16 (Change in sleeping pattern) was not loaded on either of the two factors in this solution.

Bifactor. Our three bifactor orthogonal models were derived from the three two-factor oblique solutions described above, with an added general factor specified to load on all 21 items.

Indices of reliability and support for constructs. Analyses of the BDI-II total score including item consistency and corrected item–total correlations were performed. Pearson's correlations were used to evaluate the association of the BDI-II total score with the BASIS-24 Overall, Depression/Functioning, Substance Abuse, and Psychosis scales.

ROC analyses. To evaluate the diagnostic performance of the BDI-II using the SCID-I as the reference standard, the area under the curve (AUC) was calculated for an ROC curve comparing the BDI-II total score with SCID-I major depressive episode (single episode and recurrent) diagnoses. Participants diagnosed with MDD via the SCID-I were classified as cases, and remaining participants, excluding those diagnosed with bipolar I and II and depressive disorders not otherwise specified (to minimize potential confounding effects from the MDD-like symptomology associated with these disorders), were considered controls in the analyses. This left 467 participants for the ROC analyses, with 342 (73.2%) participants diagnosed with MDD representing cases. Among the 125 comparison control participants, the most common diagnoses included alcohol dependence (29.6%; $n = 37$), substance dependence (29.6%; $n = 37$), alcohol abuse (17.6%; $n = 22$), generalized anxiety disorder (20%; $n = 16$), and substance abuse (14.4%; $n = 18$).

Sensitivity, specificity, positive and negative predictive values, and Youden's index (Youden, 1950)—the computed difference between the true and false-positive rate independent of prevalence—were calculated for each depression score to determine the optimal MDD diagnostic cutoff. The maximum Youden's index score indicates the most efficient cutoff point for optimizing both sensitivity and specificity. Although the BDI-II authors (Beck, Steer, & Brown, 1996) favored sensitivity over specificity when screening depression in clinical outpatients, for adult clinical inpatients with increased likelihood for elevated depression and multiple diagnoses at admission (Oldham et al., 1995), we considered the cutoff score offering the greatest differentiating capacity (i.e., giving equal weight to sensitivity and specificity) to be most clinically relevant.

Results

Sample Severity of Depression

The mean BDI-II total score for the 1,904 inpatients at admission was 24.93 ($SD = 13.00$), indicating moderately severe levels of depressive symptomology (Beck, Steer, & Brown, 1996). Sample means for the BASIS-24 Depression/Functioning and Overall scale scores were 2.31 ($SD = 0.98$) and 1.70 ($SD = 0.67$), respectively.

Within the SCID-I subsample ($n = 575$), the mean BDI-II total score was 25.26 ($SD = 12.70$). The subsample prevalence rate of MDD was 59.50% ($n = 342$), and participants diagnosed with MDD reported a mean BDI-II total score of 28.00 ($SD = 11.35$), the upper bound of the moderately severe range of depression (Beck, Steer, & Brown, 1996). Participants diagnosed with MDD also reported significantly higher levels of depressive symptomology than did participants diagnosed with other mental disorders ($M = 21.24$, $SD = 13.52$), $t(573) = -6.48$, $p < .001$; Cohen's $d = 0.54$.

Overview of Confirmatory Factor Analyses

To determine the optimal factor structure of the BDI-II, we first attempted to fit the unidimensional solution to the sample data. Then, we fit three previously reported two-factor oblique solutions followed by their three bifactor counterparts to the data. The bifactor solutions allowed us to test the independent contributions of a general factor and two specific factors by calculating the percentage of ECV for each tested factor. Factor intercorrelations and indices of model fit for the unidimensional, two-factor oblique, and bifactor orthogonal solutions are detailed in Table 1.

Unidimensional. The single factor solution produced poor fit to the sample data, failing to meet acceptable fit thresholds on all reported fit statistics.

Beck, Steer, and Brown (1996). The two-factor cognitive and somatic–affective oblique solution obtained from the initial BDI-II validation sample of clinical outpatients did not fit sample data according to the CFI and SRMR indices but did meet the RMSEA threshold for acceptable fit. Model factors were highly intercorrelated.

The corresponding bifactor solution produced acceptable fit to sample data across all fit statistics. As anticipated due to the large sample size, the chi-square test statistic was significant. Percentage of ECV results indicated that the general factor accounted for 84.64% of the common item variance, with the cognitive group factor accounting for 10.91% and the somatic–affective group factor accounting for 4.46% of remaining common item variance.

Osman et al. (2004). The two-factor cognitive–affective and somatic oblique solution identified using a sample of clinical adolescent inpatients did not provide acceptable fit to sample data, failing to satisfy fit criteria. The two factors were strongly intercorrelated.

Table 1
Summary of Model Fit Indices From Confirmatory Factor Analyses

Model	χ^2	AIC	Factor *r*	CFI	SRMR	RMSEA	90% CI
Unidimensional: G	3,248.091	3,372.091	—	.835	.093	.092	.089, .095
Beck, Steer, & Brown (1996): C S-A	2,246.155	2,370.155	.886	.889	.114	.075	.073, .078
Bifactor: G: C S-A	1,501.073	1,665.073	—	.928	.074	.064	.061, .067
Osman et al. (2004): C-A S	2,625.736	2,743.736	.889	.863	.109	.087	.083, .090
Bifactor: G: C-A S	1,213.230	1,371.230	—	.941	.065	.061	.058, .064
Steer et al. (2000): C N	2,377.421	2,495.421	.899	.877	.106	.082	.079, .085
Bifactor: G: C N	1,408.102	1,566.102	—	.930	.071	.066	.063, .069

Note. G = general factor; C S-A = cognitive and somatic–affective; C-A S = cognitive–affective and somatic; C N = cognitive and noncognitive; AIC = Akaike information criterion. Well-fitting model cutoff scores for fit indices: CFI (comparative fit index) = .90–.95; SRMR (standardized root-mean-square residual) ≤ .08; RMSEA (root-mean-square error of approximation) ≤ .08.

The corresponding bifactor solution demonstrated the best fit to the sample data of all solutions, improving upon Osman et al.'s poor-fitting two-factor solution on all fit indices. The general factor in this solution explained 84.86% of the common item variance, and the cognitive–affective and somatic specific factors explained 9.10% and 6.04% of the common variance, respectively.

Steer et al. (2000). The two-factor cognitive and noncognitive oblique model observed in older adult inpatients with depression provided inadequate fit to sample data, with RSMEA, CFI, and SRMR falling short of their respective cutoffs. Model factors were highly intercorrelated.

The corresponding bifactor solution for Steer et al.'s (2000) two-factor oblique solution fit the sample data well, with acceptable fit indicated by all fit statistics. In this solution, 85.12% of the common item variance was explained by the general factor, with the cognitive and noncognitive specific factors accounting for 6.34% and 8.54% of the common variance, respectively.

Comparisons of the Akaike information criterion (AIC; Akaike, 1987) values further confirmed that the bifactor orthogonal solutions rendered superior model fit to sample data versus their two-factor oblique counterparts. Comparing the AIC values, the bifactor solution based on Osman et al.'s (2004) two-factor oblique solution extracted from adolescent inpatients exhibited optimal model fit.

To sum, none of the three previously reported two-factor oblique solutions fit the data well. In contrast, all three corresponding bifactor solutions plausibly fit the data regardless of item composition of the specific factors, suggesting that the general depression factor primarily accounted for variation in the BDI-II items. Percentage of ECV findings confirmed the strength of the general factor, which contributed 84–85% of the common item variance compared to the specific factors, which explained a scant 4%–11% of the common variance. Therefore, nearly all of the common item variance in the subscales was accounted for by the general factor, supporting further analysis of the BDI-II total scale characteristics reported below but not BDI-II subscales.

Reliability indices and correlations. Within the total sample (N = 1,904), the total item coefficient alpha was .93, indicating high internal consistency for the BDI-II total score. Corrected item–total correlations ranged from .37 (Punishment feelings) to .74 (Worthlessness). Means, standard deviations, and corrected item totals for the BDI-II are provided in Table 2.

The BDI-II total score strongly correlated with the BASIS-24 Depression/Functioning subscale (r = .79, p < .001) and the BASIS-24 Overall score (r = .82, p < .001), suggestive of possible convergent validity. The relatively weaker intercorrelations between the BDI-II total score and the BASIS-24 Substance Abuse (r = .13, p < .001) and Psychosis (r = .24, p < .001) subscales were suggestive of possible discriminant validity.

ROC analyses. The AUC (95% confidence interval) of the ROC curve for the diagnostic subsample was .695 (.651–.736), indicating poor BDI-II screening performance for MDD compared to the SCID-I. ROC analyses identified an optimal BDI-II total cutoff score of 19, which possessed sensitivity of .789 and specificity of .544 (Youden's index = .333). Table 3 displays the sensitivity, specificity, positive and negative predictive values, and Youden's index for the likely optimal range of BDI-II cutoff scores.

Discussion

Present study findings support using the BDI-II total score to screen for depressive symptomology but not to generate presumptive diagnoses of MDD in adult clinical inpatients. Within our

Table 2
Descriptive Statistics for the Beck Depression Inventory–II Items

Item	*M*	*SD*	Corrected item–total correlation
Sadness	1.14	0.93	.70
Pessimism	1.19	0.96	.67
Past failure	1.38	0.98	.65
Loss of pleasure	1.44	0.94	.72
Guilty feelings	1.18	0.97	.59
Punishment feelings	0.87	1.11	.37
Self-dislike	1.61	1.05	.65
Self-criticalness	1.35	0.99	.66
Suicidal thoughts or wishes	0.47	0.65	.45
Crying	1.05	1.05	.53
Agitation	0.97	0.85	.46
Loss of interest	1.42	1.09	.73
Indecisiveness	1.37	1.08	.68
Worthlessness	1.28	1.06	.74
Loss of energy	1.32	0.88	.69
Change in sleeping pattern	1.41	0.97	.46
Irritability	0.82	0.85	.52
Changes in appetite	0.97	0.92	.43
Concentration difficulty	1.38	0.90	.65
Tiredness or fatigue	1.27	0.99	.64
Loss of interest in sex	1.05	1.11	.44

Table 3
Diagnostic Characteristics of the Beck Depression Inventory-II (BDI-II) Total Score in the Likely Optimal Range

Cutoff score	Sensitivity	Specificity	PPV	NPV	Youden's index
11	.944	.296	78.6	66.1	.240
12	.942	.352	79.9	68.7	.294
13	.924	.392	80.6	65.3	.316
14	.912	.408	80.8	63.0	.320
15	.886	.424	80.8	57.6	.310
16	.871	.432	80.8	55.1	.303
17	.839	.448	80.6	50.5	.287
18	.810	.464	80.5	47.2	.274
19	**.789**	**.544**	**82.6**	**48.6**	**.333**
20	.754	.560	82.4	45.5	.314
21	.719	.568	82.0	42.5	.287
22	.687	.584	81.9	40.6	.271
23	.658	.616	82.4	39.7	.274

Note. PPV = positive predictive value; NPV = negative predictive value. Bold indicates the optimal BDI-II total cutoff score.

large sample, on average, participants reported experiencing moderately severe levels of depression at admission, confirming the importance of screening for depression among newly admitted psychiatric inpatients.

Factor analyses indicated that bifactor BDI-II solutions specifying a general factor and two specific factors best fit our adult inpatient data, highlighting the BDI-II's concomitant uni- and multidimensional assessment of depression in adult inpatients, a trait previously identified only in other clinical and community populations (Brouwer et al., 2013; Quilty et al., 2010; Ward, 2006). Within the bifactor solutions, the strong general factor accounted for most of the common item variance, while weak specific factors contributed nominally above and beyond the general factor's influence. Collectively, this result—coupled with our failed attempts to reproduce three previously reported two-factor solutions lacking a specified general factor (Beck, Steer, & Brown, 1996; Osman et al., 2004; Steer et al., 2000) in our data—suggests that the BDI-II total score may be valid for screening depression in adult inpatients but that subscales may not produce additional clinical insights.

Subsequent reliability analyses of the BDI-II total score revealed excellent internal consistency, as evidenced by a high coefficient alpha consistent with that found in previous studies (Beck, Steer, & Brown, 1996; Steer et al., 1997, 2000). Preliminary support for the BDI-II general factor construct was demonstrated by strong correlations between the BDI-II total score and the BASIS-24 Overall and Depression/Functioning subscales, as well as by weak correlations between the BDI-II total score and the BASIS-24 Substance Abuse and Psychosis subscales.

Diagnostically, the BDI-II was unsuccessful at distinguishing adult inpatients diagnosed with MDD from inpatients diagnosed with other mental disorders, as the low AUC score generated for the BDI-II in comparison to the SCID-I marginally exceeded chance performance. Furthermore, optimal BDI-II clinical cutoff scores contained poor sensitivity–specificity balance. These subpar ROC results also precluded evaluation of established severity cutoff scores (Beck, Steer, & Brown, 1996) or the development of adult inpatient-specific severity norms, providing compelling evidence against using the BDI-II to diagnose MDD or otherwise categorize adult inpatients by depression severity.

One explanation for the BDI-II's meager diagnostic performance is that the total score may be tapping into a general factor common across multiple psychiatric conditions, perhaps related to the latent liability factors underlying certain diagnostic comorbidities (Krueger & Markon, 2006a), in addition to MDD-specific symptomology in adult inpatients. For example, Arbisi et al. (2012) found that the BDI-II performed as well as the PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993) in predicting posttraumatic stress disorder (PTSD) in military veterans, leading them to conclude that the BDI-II total score measured nonspecific distress in addition to symptoms of depression. In the current study, the extremely strong association between the BDI-II total score and the BASIS-24 Overall score, a measure of nonspecific mental health distress cutting across multiple disorders (Eisen et al., 1999), as well as moderately severe BDI-II mean total scores reported by our participants without MDD, tenders further support for the theory that the BDI-II assesses transdiagnostic generalized distress along with core features of depression in adult inpatients.

As a result, the BDI-II may be incapable of discriminating between disorders with shared underlying characteristics such as internalizing disorders (e.g., depression, anxiety, phobia) due to their significant diagnostic overlap (Krueger & Markon, 2006b), especially in adult clinical inpatients, who frequently evidence high diagnostic comorbidity (Melartin et al., 2002; Oldham et al., 1995). Instead, our data suggest that clinicians working with adult inpatients should use the BDI-II total score (and avoid using subscales) to measure their severity of presenting depression/distress but should also evaluate the BDI-II overall symptom pattern and the endorsed item content to make their clinical determinations (Beck, Steer, & Brown, 1996). Additionally, we suggest that the BDI-II total score be used in concert with other psychopathology assessments to identify significant contributions from co-occurring anxiety, psychotic, and substance use symptoms.

Several factors limit interpretation of study results. First, our sample consisted of predominantly White adults, and all participants received private psychiatric hospital care. Thus, our findings may be exclusive to this clinical subpopulation or treatment setting. Future studies should contrast the BDI-II against alternative diagnostic approaches, including other self-report depression measures such as the Patient Health Questionnaire-9 (Kroenke & Spitzer, 2002), clinical observations, and diagnostic interviews with private- and public-sector adult inpatients, to confirm our findings. Also, the multivariate nonnormality of our BDI-II data (stemming from predictably elevated severity of depression among our inpatient participants) may have influenced study results despite our best efforts to use statistical approaches robust to nonnormality. In addition, diagnostic data for the total sample (N = 1,904) could not be reported, due to the poor characterization of multiple patient diagnoses stemming from the interview-based diagnostic process that predated use of the SCID-I with our ROC analysis subsample. Finally, the presence of diagnostic comorbidity in the ROC analysis subsample, although characteristic of this clinical population (Melartin et al., 2002), may have negatively impacted ROC findings. Also, the 20% prevalence of generalized anxiety disorder among the 125 comparison control participants in the ROC subsample may have slightly confounded the BDI-II's diagnostic performance, as generalized anxiety disorder is an internalizing condition that may share a common vulnerability with MDD (Krueger & Markon, 2006a).

In closing, this large-scale psychometric investigation tentatively sanctions utilizing the BDI-II total score to screen adult clinical inpatients for depressive symptomology, yet sharply challenges its functionality as a MDD diagnostic measure. Based on study findings, the BDI-II appears to be well suited for assessing levels of depression/distress and changes in severity over time among adults requiring psychiatric hospitalization (Clapp et al., 2013; Fowler, Allen, Oldham, & Frueh, 2013) but not for diagnosing clinical depression at admission, as it may result in erroneous diagnostic conclusions leading to misdirected targeting of treatment, mitigated treatment outcomes, and impaired mental health recovery.

References

Akaike, H. (1987). Factor analysis and the AIC. *Psychometrika, 52,* 317–332. doi:10.1007/BF02294359

Allen, J. G., Frueh, B. C., Ellis, T. E., Mahoney, J. S., Oldham, J. M., Sharp, C., & Wallin, L. (2009). Integrating outcomes assessment and research into clinical care in inpatient adult psychiatric treatment. *Bulletin of the Menninger Clinic, 73,* 259–295. doi:10.1521/bumc.2009.73.4.259

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Angst, F., Stassen, H. H., Clayton, P. J., & Angst, J. (2002). Mortality of patients with mood disorders: Follow-up over 34–38 years. *Journal of Affective Disorders, 68,* 167–181. doi:10.1016/S0165-0327(01)00377-9

Arbisi, P. A., Kaler, M. E., Kehle-Forbes, S. M., Erbes, C. R., Polusny, M. A., & Thuras, P. (2012). The predictive validity of the PTSD Checklist in a nonclinical sample of combat-exposed National Guard troops. *Psychological Assessment, 24,* 1034–1040. doi:10.1037/a0028014

Arbuckle, J. L. (2012). *IBM SPSS Amos 21 user's guide.* Armonk, NY: IBM.

Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology, 20,* 112–119. doi:10.1037/0278-6133.20.2.112

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment, 67,* 588–597. doi:10.1207/s15327752jpa6703_13

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100. doi:10.1016/0272-7358(88)90050-5

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571. doi:10.1001/archpsyc.1961.01710120031004

Blatt, S. J. (2004). *Experiences of depression: Theoretical, clinical, and research perspectives.* Washington, DC: American Psychological Association. doi:10.1037/10749-000

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21,* 205–229. doi:10.1177/0049124192021002004

Brouwer, D., Meijer, R. R., & Zevalkink, J. (2013). On the factor structure of the Beck Depression Inventory-II: G is the key. *Psychological Assessment, 25,* 136–145. doi:10.1037/a0029228

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology, 41,* 193–208. doi:10.1111/j.2044-8317.1988.tb00896.x

Buckley, T. C., Parker, J. D., & Heggie, J. (2001). A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. *Journal of Substance Abuse Treatment, 20,* 197–204. doi:10.1016/S0740-5472(00)00169-0

Cameron, I. M., Cunningham, L., Crawford, J. R., Eagles, J. M., Eisen, S. V., Lawton, K., . . . Hamilton, R. J. (2007). Psychometric properties of the BASIS-24 (Behaviour and Symptom Identification Scale-Revised) mental health outcome measure. *International Journal of Psychiatry in Clinical Practice, 11,* 36–43. doi:10.1080/13651500600885531

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189–225. doi:10.1207/s15327906mbr4102_5

Clapp, J. D., Grubaugh, A. L., Allen, J. G., Mahoney, J., Oldham, J. M., Fowler, J. C., . . . Frueh, B. C. (2013). Modeling trajectory of depressive symptoms among psychiatric inpatients: A latent growth curve approach. *Journal of Clinical Psychiatry, 74,* 492–499. doi:10.4088/JCP.12m07842

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55,* 584–594. doi:10.1037/0022-006X.55.4.584

Cole, J. C., Grossman, I., Prilliman, C., & Hunsaker, E. (2003). Multimethod validation of the Beck Depression Inventory-II and Grossman-Cole Depression Inventory with an inpatient sample. *Psychological Reports, 93,* 1115–1129. doi:10.2466/pr0.2003.93.3f.1115

Costello, C. G. (1982). Social factors associated with depression: A retrospective community study. *Psychological Medicine, 12,* 329–339. doi:10.1017/S0033291700046663

Currie, S. R., Patten, S. B., Williams, J. V., Wang, J., Beck, C. A., El-Guebaly, N., . . . Maxwell, C. (2005). Comorbidity of major depression with substance use disorders. *Canadian Journal of Psychiatry, 50,* 660–666.

Dolle, K., Schulte-Körne, G., O'Leary, A. M., von Hofacker, N., Izat, Y., & Allgaier, A.-K. (2012). The Beck Depression Inventory-II in adolescent mental health patients: Cut-off scores for detecting depression and rating severity. *Psychiatry Research, 200,* 843–848. doi:10.1016/j.psychres.2012.05.011

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10,* 83–89. doi:10.1037/1040-3590.10.2.83

Eisen, S. V., Gerena, M., Ranganathan, G., Esch, D., & Idiculla, T. (2006). Reliability and validity of the BASIS-24 mental health survey for Whites, African-Americans, and Latinos. *Journal of Behavioral Health Services & Research, 33,* 304–323. doi:10.1007/s11414-006-9025-3

Eisen, S. V., Normand, S.-L., Belanger, A. J., Spiro, A., & Esch, D. (2004). The Revised Behavior and Symptom Identification Scale (BASIS-R): Reliability and validity. *Medical Care, 42,* 1230–1241. doi:10.1097/00005650-200412000-00010

Eisen, S. V., Wilcox, M., Leff, S. H., Schaefer, E., & Culhane, M. A. (1999). Assessing behavioral health outcomes in outpatient programs: Reliability and validity of the BASIS-32. *Journal of Behavioral Health Services & Research, 26,* 5–17. doi:10.1007/BF02287790

Elhai, J. D., Contractor, A. A., Biehn, T. L., Allen, J. G., Oldham, J., Ford, J. D., . . . Frueh, B. C. (2013). Changes in the Beck Depression Inventory-II's underlying symptom structure over 1 month of inpatient treatment. *Journal of Nervous and Mental Disease, 201,* 371–376. doi:10.1097/NMD.0b013e31828e1004

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling, 4,* 87–107. doi:10.1080/10705519709540063

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM–IV–TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/PSY SCREEN).* New York, NY: Biometrics Research, New York State Psychiatric Institute.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299. doi:10.1037/1040-3590.7.3.286

Fowler, J. C., Allen, J. G., Oldham, J. M., & Frueh, B. C. (2013). Exposure to interpersonal trauma, attachment insecurity, and depression severity. *Journal of Affective Disorders, 149,* 313–318. doi:10.1016/j.jad.2013.01.045

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Johnson, J., Weissman, M. M., & Klerman, G. L. (1992). Service utilization and social morbidity associated with depressive symptoms in the community. *Journal of the American Medical Association, 267,* 1478–1483. doi:10.1001/jama.1992.03480110054033

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association, 289,* 3095–3105. doi:10.1001/jama.289.23.3095

Kessler, R. C., Nelson, C. B., McGonagle, K. A., Liu, J., Swartz, M., & Blazer, D. G. (1996). Comorbidity of the *DSM–III–R* major depressive disorder in the general population: Results from the US National Comorbidity Survey. *British Journal of Psychiatry, 30,* 17–30.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Krefetz, D. G., Steer, R. A., Gulab, N. A., & Beck, A. T. (2002). Convergent validity of the Beck Depression Inventory-II with the Reynolds Adolescent Depression Scale in psychiatric inpatients. *Journal of Personality Assessment, 78,* 451–460. doi:10.1207/S15327752JPA7803_05

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32,* 509–515.

Krueger, R. F., & Markon, K. E. (2006a). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2,* 111–133. doi:10.1146/annurev.clinpsy.2.022305.095213

Krueger, R. F., & Markon, K. E. (2006b). Understanding psychopathology: Melding behavior genetics, personality, and quantitative psychology to develop an empirically based model. *Current Directions in Psychological Science, 15,* 113–117. doi:10.1111/j.0963-7214.2006.00418.x

Lindsay, W. R., & Skene, D. D. (2007). The Beck Depression Inventory II and the Beck Anxiety Inventory in people with intellectual disabilities: Factor analyses and group data. *Journal of Applied Research in Intellectual Disabilities, 20,* 401–408. doi:10.1111/j.1468-3148.2007.00380.x

MedCalc Software. (2013). *MedCalc statistics for biomedical research: Software manual.* Ostend, Belgium: Author.

Melartin, T. K., Rytsala, H. J., Leskela, U. S., Lestela-Mielonen, P. S., Sokero, T. P., & Isometsa, E. T. (2002). Current comorbidity of psychiatric disorders among *DSM–IV* major depressive disorder patients in psychiatric care in the Vantaa Depression Study. *Journal of Clinical Psychiatry, 63,* 126–134. doi:10.4088/JCP.v63n0207

Oldham, J. M., Skodol, A. E., Kellman, H. D., Hyler, S. E., Doidge, N., Rosnick, L., & Gallaher, P. E. (1995). Comorbidity of axis I and axis II disorders. *American Journal of Psychiatry, 152,* 571–578.

Osman, A., Barrios, F. X., Gutierrez, P. M., Williams, J. E., & Bailey, J. (2008). Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. *Journal of Clinical Psychology, 64,* 83–102. doi:10.1002/jclp.20433

Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., & Bagge, C. L. (2004). Reliability and validity of the Beck Depression Inventory-II with adolescent psychiatric inpatients. *Psychological Assessment, 16,* 120–132. doi:10.1037/1040-3590.16.2.120

Quilty, L. C., Zhang, A. Z., & Bagby, R. M. (2010). The latent symptom structure of the Beck Depression Inventory-II in outpatients with major depression. *Psychological Assessment, 22,* 603–608. doi:10.1037/a0019698

Reise, S. P., Moore, T. N., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scores. *Journal of Personality Assessment, 92,* 544–559. doi:10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcome measures. *Quality of Life Research, 16,* 19–31. doi:10.1007/s11136-007-9183-7

Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis, 10,* 235–249. doi:10.1016/0167-9473(90)90004-2

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research, 99,* 323–338. doi:10.3200/JOER.99.6.323-338

Simon, G. E., & VonKorff, M. (1998). Suicide mortality among patients treated for depression in an insured population. *American Journal of Epidemiology, 147,* 155–160. doi:10.1093/oxfordjournals.aje.a009428

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1997). Further evidence for the construct validity of the Beck Depression Inventory-II with psychiatric outpatients. *Psychological Reports, 80,* 443–446. doi:10.2466/pr0.1997.80.2.443

Steer, R. A., Clark, D. A., Beck, A. T., & Ranieri, W. F. (1999). Common and specific dimensions of self-reported anxiety and depression: The BDI-II versus the BDI-IA. *Behaviour Research and Therapy, 37,* 183–190. doi:10.1016/S0005-7967(98)00087-4

Steer, R. A., Rissmiller, D. J., & Beck, A. T. (2000). Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behaviour Research and Therapy, 38,* 311–318. doi:10.1016/S0005-7967(99)00068-6

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42,* 893–898. doi:10.1016/j.paid.2006.09.017

Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-Second Edition in a sample of college students. *Depression and Anxiety, 19,* 187–189. doi:10.1002/da.20002

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. doi:10.1177/1073191107311261

Ward, L. C. (2006). Comparison of factor structure models for the Beck Depression Inventory-II. *Psychological Assessment, 18,* 81–88. doi:10.1037/1040-3590.18.1.81

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD Checklist (PCL): Reliability, validity and diagnostic utility*. Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology, 106,* 203–214.

Wells, K. B., Stewart, A., Hays, R. D., Burnam, A., Rogers, W., Daniels, M., . . . Ware, J. (1989). The functioning and well-being of depressed patients: Results from the Medical Outcomes Study. *Journal of the American Medical Association, 262,* 914–919. doi:10.1001/jama.1989.03430070062031

Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56,* 545–551. doi:10.1002/(SICI)1097-4679(200004)56:4<545::AID-JCLP7>3.0.CO;2-U

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32–35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Received August 14, 2013
Revision received March 17, 2014
Accepted March 25, 2014 ■