Health Psychology
2001, Vol. 20, No. 2, 112–119

0278-6133/01/$5.00 DOI: 10.1037//0278-6133.20.2.112

Psychometric Evaluation of the Beck Depression Inventory-II With Primary Care Medical Patients

Randolph C. Arnau, Mary W. Meagher, Margaret P. Norris, and Rachel Bramson
Texas A&M University

This study evaluated the psychometric characteristics of the Beck Depression Inventory-II (BDI-II; A. T. Beck, R. A. Steer, & G. K. Brown, 1996) in a primary care medical setting. A principal-components analysis with Promax rotation indicated the presence of 2 correlated factors, Somatic-Affective and Cognitive, which explained 53.5% of the variance. A hierarchical, second-order analysis indicated that all items tap into a second-order construct of depression. Evidence for convergent validity was provided by predicted relationships with subscales from the Short-Form General Health Survey (SF-20; A. L. Stewart, R. D. Hayes, & J. E. Ware, 1988). A receiver operating characteristic analysis demonstrated criterion-related validity: BDI-II scores predicted a diagnosis of major depressive disorder (MDD), as determined by the Primary Care Evaluation of Mental Disorders (PRIME-MD) Patient Health Questionnaire (PHQ). This study demonstrated that the BDI-II yields reliable, internally consistent, and valid scores in a primary care medical setting, suggesting that use of the BDI-II in this setting may improve detection and treatment of depression in these medical patients.

Key words: Beck Depression Inventory-II, psychometrics, primary care, validity, reliability, depression

Depression is one of the most common psychological disorders seen in primary care medical patients, with prevalence rates as high as 23% (Coyne, Fechner-Bates, & Schwenk, 1994; Regier et al., 1993; Zung, 1990). The detection of depression in primary care patients is especially important because 50–75% of individuals seeking treatment for a depressive disorder present to their physician, whereas only 16–23% present to a mental health practitioner (Depression Guideline Panel, 1993; Muñoz, Hollon, McGrath, Rehm, & VandenBos, 1994). Unfortunately, because depression often is not detected in primary care settings, over half of the patients in these settings with major depressive disorder (MDD) remain untreated (Coyne, Fechner-Bates, & Schwenk, 1995; Montono, 1994; Vasquez-Barquero, Herran, & Artal, 1997). Patients with untreated depression exhibit significant functional impairments and higher morbidity and mortality than average, and tend to be high utilizers of medical services (Eisenberg, 1992; Wells et al., 1989). To increase rates of detection and treatment of depression in primary care settings, primary care providers are turning to screening instruments (Katon et al., 1997) such as the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996). However, despite its widespread use, relatively little is known about the psychometric properties of the BDI-II in primary care patients.

The BDI-II is a revised version of the 21-item Beck Depression Inventory (BDI; Beck & Steer, 1993), which assesses the severity of depression in adults and adolescents. The BDI-II is an update of the original BDI, which was altered to correspond to criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV;* American Psychiatric Association, 1994) for major depressive disorder and to improve the content validity of the instrument. For example, the revised instrument includes modifications to 17 responses, including options for both increases and decreases in appetite, weight, and sleep. In addition, four items were dropped (body image change, work difficulty, weight loss, and somatic preoccupation) and replaced by four new items (agitation, worthlessness, loss of energy, and concentration difficulty). The time frame for responses was lengthened from 1 week to 2 weeks to be consistent with the *DSM-IV* temporal criterion for a major depressive episode.

The BDI-II has been validated with college students (Beck et al., 1996; Dozois, Dobson, & Ahnberg, 1998; Osman et al., 1997), adult psychiatric outpatients (Beck et al., 1996; Steer, Ball, Ranieri, & Beck, 1999), and adolescent psychiatric outpatients (Steer, Geetha, Ranieri, & Beck, 1998). Although the BDI-II demonstrated excellent test–retest reliability, high internal consistency, and moderate to high convergent validity, the factor structure varied across studies. Using exploratory factor analysis on a college-student sample, Beck et al. (1996) obtained a two-factor solution involving a Cognitive-Affective factor and a Somatic factor. The same two-factor solution was found for another college-student sample using both exploratory and confirmatory factor analyses (Dozois, Dobson, & Ahnberg, 1998). In contrast, Osman et al. (1997) reported a three-factor model corresponding to Negative Attitude, Performance Difficulty, and Somatic clusters in college students. In adult psychiatric outpatients, a two-factor structure corresponding to a Somatic-Affective factor and Cogni-

Randolph C. Arnau, Mary W. Meagher, and Margaret P. Norris, Department of Psychology, Texas A&M University; Rachel Bramson, College of Medicine, Texas A&M University.

This research was funded by grants from Scott & White Hospital, Temple, Texas to Mary W. Meagher and from the Bush School of Government and Public Service, College Station, Texas. Portions of this article were presented at the 107th Annual Convention of the American Psychological Association, Boston, Massachusetts, August 24, 1999. We express our appreciation to Jack Bodden for his help in facilitating this project.

Correspondence concerning this article should be addressed to Mary W. Meagher, Department of Psychology, Texas A&M University, College Station, Texas 77843-4235. Electronic mail may be sent to mwm@psyc.tamu.edu.

tive factor has been consistently obtained with both exploratory and confirmatory factor analyses (Beck et al., 1996; Steer et al., 1999). Moreover, confirmatory factor analyses of clinically depressed adult outpatients suggest that the BDI-II comprises one second-order Depression factor that can be further divided into two first-order factors (Somatic-Affective and Cognitive factors; Steer et al., 1999).

Similarly, Steer et al. (1998) found a second-order factor of Depression in a sample of adolescent psychiatric outpatients, but the structure differed from adult outpatients in that three first-order factors emerged, rather than the two that have been found in adult outpatient samples. However, two of the three factors were comparable to the Somatic-Affective and Cognitive factors found in adult outpatient samples; the third factor was composed of only three items and was deemed ungeneralizable.

Because no studies, to our knowledge, have investigated the psychometric characteristics of the BDI-II in primary care medical patients, such an investigation was the purpose of the present study. To assess reliability, we evaluated the internal consistency of item scores and item-total correlations. Validity was assessed by examining: (a) the first- and second-order factor structure using exploratory and hierarchical factor analysis, (b) convergent and discriminant validity with the Medical Outcomes Study Short-Form General Health Survey (SF-20; Stewart, Hayes, & Ware, 1988) and diagnoses of MDD and (c) receiver operating characteristics analyses to determine criterion-related validity for predicting MDD.

This study also examined the validity of somatic symptoms (e.g., disturbed eating and sleeping) as indicators of depression in primary care patients. In a sample of psychiatric outpatients, loss of energy, sleep disturbance, and appetite disturbance were the best predictors of MDD (Buchwald & Rudick-Davis, 1993). However, some have proposed that somatic symptoms of depression may lose diagnostic utility in a medical setting because somatic complaints may be confounded with physical illness (see Clark, Cavanaugh, & Gibbons, 1983). Empirical studies have consistently demonstrated that for elderly medical patients somatic symptoms are, indeed, good indicators of depression (Koenig, Cohen, Blazer, Krishnan, & Sibert, 1993; Norris, Snow-Turek, & Blankenship, 1995; Norris & Woehr, 1998). Therefore, we expected that the diagnostic utility of somatic symptoms would also be demonstrated in younger adult primary care patients.

Method

Study Setting

The setting for this study was a large staff-model health maintenance organization (HMO) in a suburban community in southeast Texas. Participants were recruited from the family medicine clinic, which is staffed by 15 board-certified family physicians. Approximately 60% of the patients presenting to this clinic subscribed to the health plan of the staff-model HMO, and the remaining approximately 40% either subscribed to other health plans or were self-pay patients.

Recruitment of Participants

All patients over age 18 who presented to the clinic for an appointment with their primary care physician were eligible. Recruitment was two-staged. First, eligible patients were contacted by telephone the day before their appointment. They were informed that a study was being conducted involving screening for "stress-related problems," and they were given a brief, standardized statement of the purpose and procedures. In the second stage of recruitment, patients who previously agreed to participate were greeted by a researcher when they arrived for their appointment. The researcher provided complete details, answered questions, and distributed the informed consent form and questionnaire packet.

A total of 1,486 patients were invited to participate; 759 (51%) declined. Because these data were collected to address several distinct investigations, the BDI-II was administered to a subset of only 340 patients.

Procedure

While waiting for their appointments with their physicians, patients gave written informed consent and completed the BDI-II, the Primary Care Evaluation of Mental Disorders (PRIME-MD) Patient Health Questionnaire (PHQ; Spitzer, Kroenke, & Williams, 1999), and the SF-20 as part of a larger battery of instruments. Following their appointments, patients returned completed questionnaires to a masters-level graduate student who scored the screening instruments and provided written feedback to the patients and their physicians.

Participants

The sample for the present study consisted of 340 primary care patients (68.8% female), 7 of whom were excluded because of missing responses. Participants' ages ranged from 18 to 54 years (M = 36.5, SD = 10.1). Other available sample demographics are summarized in Table 1. Data

Table 1
Sample Demographics

Variable	*n*	%
Sex		
Female	229	68.8
Male	104	31.2
Race/ethnicity		
Caucasian	231	69.4
African American	23	6.9
Hispanic	22	6.6
Asian American/Pacific Islander	4	1.2
Other	7	2.1
Unreported	46	13.8
Education		
Some high school	12	3.6
High school or GED test	62	16.6
Some college	82	24.6
College degree	50	15.0
Some graduate school	19	5.7
Graduate degree	60	18.0
Unreported	48	14.4
Household income		
Under $10,000	21	6.3
$10–20,000	36	10.8
$21–30,000	60	18.0
$31–50,000	80	24.0
$50–80,000	61	18.3
Over $80,000	21	6.3
Unreported	54	16.2
Marital status		
Single	56	16.8
Married	208	62.5
Separated	3	0.9
Divorced	20	6.0
Widowed	2	0.6
Unreported	44	13.2

Note. GED = general education development.

collected for older adults (from 54 to 86 years of age) were not included in the present study.[1]

Because of the number of patients who declined to participate, we evaluated whether participants and decliners differed on several dimensions. Some demographic data (gender and age) and health and medical utilization data (total number of diagnoses received in the past year and total number of visits to a doctor in the past year) were available for 210 of the decliners. Analyses of these data indicated that there were no statistically significant differences between participants and decliners for any of these variables.

Measures

BDI-II. The BDI-II[2] is a 21-item self-report measure of the severity of depressive symptomatology. Each of the 21 items is rated on a 4-point scale ranging from 0–3. The ratings are summed, yielding a total score that can range from 0–63.

SF-20. We assessed self-reported health and functioning using the SF-20, which is a 20-item self-report instrument that measures the effects of health on physical and role functioning, bodily pain, general health perceptions, and mental health. Lower scores reflect poorer perceived health and functioning.

PHQ. The PHQ is a self-report instrument designed to be used in primary care settings to diagnose a number of *DSM-IV* disorders, including MDD, panic disorder, and bulimia nervosa. For the purposes of the present study, the depression module of the PHQ was used to identify participants with MDD.

The PHQ is a revised version of the PRIME-MD Patient Problem Questionnaire (PPQ), which was developed as a structured interview. The original PPQ was validated in a study involving 1,000 primary care patients and was subsequently revised into the self-report version called the PHQ, which was intended to increase the clinical and research utility of the original PPQ.

The diagnostic accuracy of the PHQ was recently assessed with a sample of 550 primary care patients. Within 48 hr of administration of the PHQ, a mental health professional, who had not been informed of the PHQ results, contacted patients and determined diagnoses using a structured interview. This analysis demonstrated a 93% agreement between a PHQ diagnosis of MDD and diagnosis by the mental health professional (Spitzer, Kroenke, & Williams, 1999).

Results

Prevalence and Severity of Depression Symptoms

The average BDI-II score was 8.74 (SD = 9.7). Scores of 14 or higher, suggesting at least a mild level of depression (Beck et al., 1996), were observed in 23.2% of the sample.

Socioeconomic Status and Gender Differences

As expected, there was a small relationship between BDI-II scores and socioeconomic status variables. Specifically, BDI-II scores were inversely related to level of education ($r = -.24$, $p < .01$) and household income ($r = -.20$, $p < .01$). In addition, there was an expected gender difference in BDI-II scores. The mean BDI-II score for women, $M = 9.88$, $SD = 10.33$, was higher than the mean score for men, $M = 5.82$, $SD = 6.91$; $t(338) = 3.70$, $p < .001$.

Reliability

Internal consistency. The internal consistency of BDI-II responses was excellent, with an alpha coefficient of .94.

Item-total correlations. The item Ms, SDs, and corrected item-total correlations are presented in Table 2. The item-total correlations ranged from .54 to .74, indicating good internal consistency.

Factorial Validity

First-order analysis. Because little factor-analytic research on the original BDI has been conducted in a medical setting and none has been conducted on the BDI-II in a medical setting, the factor structure of BDI-II scores was examined using exploratory factor analysis. Although the factor structure was expected to be similar to that of the BDI and the BDI-II in nonmedical settings, no explicit structure was hypothesized a priori for this sample.

A principal–components analysis of the BDI-II item correlations was conducted. We based the decision about the number of factors to retain on (a) parallel analysis (Horn, 1965) and (b) the interpretability and theoretical salience of the rotated factors. Parallel analysis has consistently been shown to be superior to other factor retention rules in terms of extracting the correct number of factors in Monte Carlo studies (Crawford & Koopman, 1973; Zwick & Velicer, 1986). In the present study, parallel analysis indicated that two factors should be retained. Furthermore, when the two factors were rotated, they were theoretically salient and accounted for a total of 53.5% of the variance.

Because both factors are theorized to tap into depression, we hypothesized that the factors would be correlated. Therefore, we subjected the two factors to a Promax rotation (Hendrickson & White, 1964), which allows correlated factors and generally yields good simple structure (Gorsuch, 1983).

As expected, the rotated factors were correlated ($r = .70$). The factor-pattern coefficients are presented in Table 3, along with the communalities (h^2) of the measured variables. Factor pattern coefficients of .40 or greater were considered salient (Stevens, 1996, p. 372). The first factor was labeled Somatic-Affective, given the salient pattern and structure coefficients for measured variables making up that component, such as tiredness, sleep problems, problems with appetite, sadness, and loss of pleasure. The second factor was labeled Cognitive, given the salient pattern and structure coefficients for the measured variables making up that component, such as suicidality, pessimism, worthlessness, and guilt. One item, self-criticalness, did not make a substantial contribution to either factor, with a pattern coefficient of .37 for the Somatic-Affective component, and .32 for the Cognitive component.

Whenever factors are correlated, structure coefficients (correlations of the measured variables with the extracted components) are also important aids to interpretation (Thompson, 1997; Thompson & Borrello, 1985). Therefore, the structure coefficients are presented in Table 4. The large structure coefficients for all measured variables on both components are consistent with the high correlation between the rotated components. It is also important to note that the structure coefficients for the self-criticalness item were large, indicating this item correlates with both first-order factors and thus is a relevant symptom of depression, despite the low pattern coefficients.

[1] These data were analyzed and reported separately to examine different hypotheses than those tested in the present study (see Norris, Meagher, & Arnau, 1999).

[2] The BDI-II can be ordered from The Psychological Corporation, P.O. Box 839954, San Antonio, TX 78282-3954.

Table 2
Means, Standard Deviations, and Corrected Item-Total Correlations of the BDI-II

Item	*M*	*SD*	r_{tot}
Sadness	.28	.54	.68
Pessimism	.35	.61	.66
Past Failure	.41	.70	.69
Loss of Pleasure	.44	.68	.69
Guilty Feelings	.40	.59	.62
Punishment Feelings	.23	.65	.54
Self-Dislike	.43	.74	.74
Self-Criticalness	.41	.69	.62
Suicidal Thoughts	.08	.31	.54
Crying	.32	.72	.67
Agitation	.41	.61	.58
Loss of Interest	.40	.67	.73
Indecisiveness	.31	.68	.73
Worthlessness	.22	.56	.68
Loss of Energy	.70	.67	.64
Changes in Sleeping	.77	.95	.58
Irritability	.48	.78	.67
Changes in Appetite	.53	.83	.60
Concentration Difficulty	.46	.67	.68
Tiredness	.65	.72	.67
Loss of Interest in Sex	.48	.80	.57

Note. BDI-II = Beck Depression Inventory-II.

Second-order analysis. The correlation of the rotated factors (r = .70) implied a higher level of conceptualization, which should correspond to a second-order factor of depression. For the second-order analysis, a principal-components analysis of the first-order factor correlation matrix (in this case, a single correlation of .70) was conducted. One second-order factor was extracted, yielding pattern coefficients of .92 for both of the first-order factors loading onto the second-order factor.

Table 3
Rotated Factor Pattern Matrix for the BDI-II

	Factor		
Item	1	2	h^2
Sadness	**.474**	.323	54.3
Pessimism	−.007	**.784**	60.7
Past Failure	.055	**.761**	64.0
Loss of Pleasure	**.464**	.341	55.2
Guilty Feelings	.247	**.451**	41.9
Punishment Feelings	−.140	**.802**	50.7
Self-Dislike	.367	**.466**	59.0
Self-Criticalness	.373	.317	40.4
Suicidal Thoughts	−.299	**.936**	57.6
Crying	.235	**.516**	49.0
Agitation	**.670**	.001	45.0
Loss of Interest	**.517**	.321	60.0
Indecisiveness	**.589**	.217	57.1
Worthlessness	.062	**.741**	61.7
Loss of Energy	**.861**	−.155	58.0
Changes in Sleeping	**.744**	−.105	45.6
Irritability	**.650**	.119	54.5
Changes in Appetite	**.658**	−.005	42.9
Concentration Difficulty	**.694**	.076	56.0
Tiredness	**.942**	−.212	65.4
Loss of Interest in Sex	**.683**	−.024	44.4

Note. Pattern coefficients with values of .40 or greater are in boldface. Factor 1 = Somatic-Affective, Factor 2 = Cognitive. h^2 = Communalities of the measured variables; figures in this column represent %s. BDI-II = Beck Depression Inventory-II.

Table 4
Structure Matrix for the BDI-II

	Factor	
Item	1	2
Sadness	.699	.653
Pessimism	.538	.779
Past Failure	.584	.799
Loss of Pleasure	.701	.664
Guilty Feelings	.561	.623
Punishment Feelings	.418	.705
Self-Dislike	.691	.721
Self-Criticalness	.594	.576
Suicidal Thoughts	.352	.728
Crying	.594	.679
Agitation	.671	.467
Loss of Interest	.740	.680
Indecisiveness	.740	.627
Worthlessness	.578	.784
Loss of Energy	.753	.444
Changes in Sleeping	.671	.412
Irritability	.733	.572
Changes in Appetite	.655	.453
Concentration Difficulty	.747	.559
Tiredness	.794	.443
Loss of Interest in Sex	.666	.451

Note. BDI-II = Beck Depression Inventory-II. Factor 1 = Somatic-Affective, Factor 2 = Cognitive.

As an aid to interpretation of the second-order analysis, a Schmid–Leiman orthogonalized solution (Schmid & Leiman, 1957) was computed, as recommended by Gorsuch (1983). The Schmid–Leiman transformation yields a solution in which (a) the second-order factors are expressed in terms of the measured variables (rather than in terms of the first-order factors) and (b) the common variance accounted for by the second-order factors is residualized from the first-order factors. In other words, the pattern coefficients of the variables for the first-order factors represent the unique variance that is accounted for by the first-order factors and is not accounted for by the second-order factors (see Arnau, 1998; Gorsuch, 1983; Thompson, 1990; Thompson & Borrello, 1992).

The Schmid–Leiman solution for the present analysis was computed using Interactive Matrix Language (IML; SAS Institute, 1990) and is presented in Table 5. All of the measured variables from the BDI-II contributed a noteworthy amount of variance to the second-order depression factor, with pattern coefficients ranging from .59 to .77.

It is also apparent from Table 5 that very little variance in the first-order factors is not accounted for by the second-order factor. This is evidenced from the magnitudes of the pattern coefficients in the third and fourth columns of Table 5, which represent the unique variance still accounted for by the first-order factors after being residualized of the variance captured by the second-order factor. None of the first-order coefficients retained salience after being residualized of the variance accounted for by the second-order factor.

Table 5
Schmid–Leiman Orthogonalized Solution

Item	Factor		
	Depression	1	2
Sadness	**.734**	.185	.126
Pessimism	**.715**	−.003	.306
Past Failure	**.751**	.021	.297
Loss of Pleasure	**.741**	.181	.133
Guilty Feelings	**.643**	.096	.176
Punishment Feelings	**.610**	−.055	.313
Self-Dislike	**.767**	.143	.182
Self-Criticalness	**.635**	.145	.124
Suicidal Thoughts	**.587**	−.117	.365
Crying	**.692**	.092	.201
Agitation	**.618**	.261	.000
Loss of Interest	**.772**	.202	.125
Indecisiveness	**.742**	.230	.085
Worthlessness	**.739**	.024	.289
Loss of Energy	**.650**	.336	−.060
Changes in Sleeping	**.588**	.290	−.041
Irritability	**.708**	.253	.046
Changes in Appetite	**.602**	.257	−.002
Concentration Difficulty	**.709**	.271	.029
Tiredness	**.672**	.367	−.083
Loss of Interest in Sex	**.607**	.266	−.009

Note. The Depression column represents the second-order factor. The Factor 1 and Factor 2 columns represent the first-order solution, based on variance orthogonal to the second-order factor (Gorsuch, 1983). Coefficients with an absolute value of .40 or greater are in boldface. Factor 1 = Somatic-Affective, Factor 2 = Cognitive. Communalities (h^2) for the second-order solution are the same as those of the first-order solution presented in Table 2.

Convergent Validity

The convergent validity of the BDI-II scores was examined through correlations of BDI-II total scores and factor scores with SF-20 subscale scores, including the Mental Health, Perception of Overall Health, Pain, Physical Functioning, and Role Functioning subscales. The correlations between BDI-II total scores and factor scores with the SF-20 subscales are presented in Table 6. When interpreting these results, it is important to note that higher scores on the SF-20 indicate greater functioning and positive health perceptions, so negative relationships between these scales and BDI-II scores indicate higher depression severity associated with lower functioning and lower perceptions of health. The correlations between the total BDI-II scores and the SF-20 subscales ranged from −.19 for the Physical Functioning subscale to −.65 for the Mental Health subscale. As expected, the BDI-II total and factor scores correlated more strongly with the Mental Health subscale than with any of the other SF-20 subscales.

Criterion-Related Validity

Because the BDI-II purportedly measures severity of depressive symptomatology, there should be a difference in mean BDI-II scores between a group of patients with a diagnosis of MDD versus a group with no depressive disorder diagnosis. As mentioned previously, *DSM-IV* diagnostic criteria were assessed with the PHQ. Using this criterion, 31 patients obtained a *DSM-IV* diagnosis of MDD and 304 patients did not. The mean BDI-II scores differed substantially across the two groups, with the nondepressive diagnosis group averaging 6.7 (SD = 7.1) and the depressive diagnosis group averaging 28.0 (SD = 9.7), a difference that is noteworthy and statistically significant; $t(333) = 15.3$, $p < .001$.

Receiver Operating Characteristics (ROC) Analysis

Given the mean BDI-II score difference between patients with and without MDD, it was deemed useful to evaluate how well BDI-II scores predict the presence of MDD. When used as a screening measure, the BDI-II could alert primary care providers that a patient should receive further evaluation for a depressive disorder.

ROC analysis is specifically designed to provide an index of the diagnostic accuracy of a screening instrument (Rey, Morris-Yates, & Stanislaw, 1992). The results of the analysis can also be used to determine an optimal cutoff score. For any given cutoff score, there are several indices of diagnostic efficacy. The true positive rate (sensitivity) refers to the proportion of individuals with the disorder who are identified as such by the screen (i.e., score above the cutoff), whereas the false positive rate is the proportion of those scoring above the cutoff who actually do not have the disorder. The true negative rate (specificity) is the proportion of individuals without the disorder who are identified as such by the screen (i.e., score below the cutoff), and the false negative rate is the proportion of those scoring below the cutoff who actually have the disorder.

Two other important indices derived from an ROC analysis are positive predictive power (PPP) and negative predictive power (NPP). PPP is the ratio of individuals screening positive who actually do have the disorder to the total number of individuals who screen positive (true positives + false positives). NPP is the ratio of individuals who screen negative who do not have the disorder to the total number of individuals who screen negative (true negatives + false negatives).

In an ROC analysis, the sensitivity and specificity associated with every possible cutoff score are calculated and plotted, with sensitivity on the x axis and specificity on the y axis. The area under the curve (AUC) is an index of the amount of diagnostic information provided by an instrument (see Hanley & McNeil, 1982; Swets, 1988) and can range from 0.0 (no diagnostic information) to 1.0 (perfect diagnostic accuracy). The diagonal along the middle of the curve represents the diagnostic information that

Table 6
Convergent Validity Coefficients of the BDI-II With SF-20 Subscales

SF-20 subscale	BDI-II scores		
	Total score	Factor 1	Factor 2
Mental Health	−.65	−.61	−.59
Health Perceptions	−.42	−.43	−.34
Physical Pain	−.24	−.24	−.21
Physical Functioning	−.19	−.18	−.18
Role Functioning	−.35	−.35	−.31

Note. Higher scores on SF-20 subscales indicate better health and greater functioning. BDI-II = Beck Depression Inventory-II. Factor 1 = Somatic-Affective, Factor 2 = Cognitive. All correlations are statistically significant at the $p < .01$ level.

could be expected by chance, and represents an AUC of .50. An AUC of .80 or higher indicates a useful screening instrument (Holmes, 1998).

The computer program Simstat (Provalis Research, 1996) was used to conduct an ROC analysis with BDI-II scores predicting the diagnosis of MDD (determined by the PHQ, as described previously). The AUC was .96, indicating that the BDI-II exhibits excellent performance as a screen for MDD.

The ROC analysis indicated that using a BDI-II cutoff score of 18 yielded the best balance between sensitivity and specificity. Table 7 depicts the percentages of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) resulting from this cutoff score in the present sample. This classification distribution yielded a sensitivity of 94% and a specificity of 92%, with an overall correct classification rate of 92%. In addition, this yielded a PPP of 54% and an NPP of 99%. These values illustrate the fact that although false positive results were produced, almost all of those who actually had MDD were identified as such (i.e., scored above the cutoff). There were very few FN (i.e., patients with MDD who were not identified as such).

Although 18 is the cutoff recommended for the best balance between sensitivity and specificity, one might opt for either higher sensitivity or higher specificity. The ROC analyses indicated that if higher sensitivity is desired (fewer FN), a cutoff of 10 is recommended, which yields a sensitivity of 100% and specificity of 70%. If higher specificity is desired (fewer FP), a cutoff of 25 is recommended, which yields a sensitivity of 58% and specificity of 97%.

Discussion

This study evaluated the reliability and validity of BDI-II scores in a primary care medical setting. The findings indicate that the BDI-II yields reliable and valid scores for assessing depression in this setting. With regard to reliability, BDI-II scores demonstrated excellent internal consistency. In addition, corrected item-total correlations ranged from .54 to .74, which are somewhat higher than those reported by Osman et al. (1997; range = .44 to .65) and Dozois et al. (1998; range = .41 to .62).

This study also yielded strong evidence for the factorial validity of the BDI-II in a primary care setting. Principal-components analysis suggested that two factors (Somatic-Affective and Cognitive) summarized the data parsimoniously. Although a few individual items loaded differently, the two extracted factors were the same factors found for adult psychiatric outpatients (Beck et al., 1996), and comparable to the first two factors found using adolescent psychiatric outpatients (Steer et al., 1998). Conversely, the factor structure we found for primary care medical patients differed from that of college students, which have typically yielded a Cognitive-Affective factor and a Somatic factor (Beck et al., 1996; Dozois et al., 1998). Therefore, it seems that the structure of depression in primary care medical patients is similar to that of both adult and adolescent psychiatric outpatients.

As mentioned previously, a few items in our sample demonstrated different factor patterns than those found for a psychiatric outpatient sample reported by Beck et al. (1996). In our sample, sadness loaded most highly on the Somatic-Affective factor, whereas this item loaded on the Cognitive factor in the psychiatric outpatient sample. However, sadness is probably best conceptualized as consisting of affective and cognitive components, and thus might be expected to contribute to an affective or a cognitive factor.

Consistent with our findings, crying contributed to the Cognitive factor in the psychiatric sample studied by Steer et al. (1999), although crying was mainly associated with the Somatic-Affective factor in the Beck et al. (1996) sample. These differences are consistent with the hypothesis of Beck et al. (1996) that affective symptoms may be located in different dimensions depending on the background and diagnostic composition of the sample under study.

The present study by presenting results of a second-order factor analysis in a primary care medical setting, expands the literature on the BDI-II. The Schmid-Leiman orthogonalized solution indicated one second-order factor made up of strong contributions by every item on the BDI-II. This gave stronger evidence for the construct validity of the BDI-II than could be provided by a first-order analysis alone because the second-order analysis produced a more accurate model of what the BDI-II is intended to measure (i.e., a single construct of depression). Specifically, the second-order analysis indicated that the instrument measures distinct, albeit related, factors of depression that both tap into a higher-order factor that can be called depression. Furthermore, the first-order factors account for only a small amount of variance after being residualized of the variance accounted for by the higher-order factor of depression.

The strong correlations of BDI-II scores with the Mental Health subscale of the SF-20 provided evidence for the convergent validity of BDI-II scores. In addition, BDI-II scores were moderately correlated with SF-20 subscales corresponding to general perceptions of health and social-role functioning. Consistent with other studies, depression was associated with physical functioning and physical pain (Romano & Turner, 1985; Williamson & Schulz, 1992). In addition, noteworthy and statistically significant differences in total BDI-II scores across groups of patients who were and were not diagnosed with MDD demonstrated discriminant validity for the BDI-II.

The ROC analysis indicated that a cutoff score of 18 yielded the best balance between sensitivity and specificity for predicting MDD in this setting, which led to an overall rate of correct classification of 92%. These results are particularly positive given the widespread call for psychological screening in primary care medical settings due to the underdetection of depression in such settings (Katon et al., 1997; Montono, 1994). Results from the ROC analysis indicate that the BDI-II is a very sensitive and moderately specific screen for depression in a primary care medical setting. Because the BDI-II takes only 5–10 min to complete and is easily scored, it could be integrated into a primary care setting without slowing patient flow. Patients scoring above a

Table 7
Contingency Table Obtained When Using a Cutoff Score of 18 on the BDI-II

	BDI-II score	
Diagnosis	Score > 18	Score < 18
MDD	n = 29 (true positives)	n = 2 (false negatives)
No MDD	n = 25 (false positives)	n = 279 (true negatives)

Note. MDD = major depressive disorder.

cutoff score could be flagged for follow-up by the physician or mental health provider.

It is important to note that the PPP (.54) indicated that the false positive rate is almost as high as the true positive rate. In this sample, for patients scoring at or above the cutoff of 18, there was a 54% chance that the patient had MDD. However, a lower PPP than NPP is not inconsistent with the main goal of screening in primary care settings, which is to increase the rate of detection. The results of this analysis indicate that the BDI-II missed only 1% of patients with MDD; in contrast, previous studies have shown that without screening, sometimes over half the patients with MDD in primary care settings go untreated (Coyne et al., 1995; Montono, 1994; Vasquez-Barquero et al., 1997). The practical implication of relatively lower PPP is that patients scoring positive on screens should be flagged for further evaluation by a physician or mental health practitioner to determine if they should be treated for depression. The present study suggests that the BDI-II is an excellent screen for identifying patients who may need treatment for MDD.

Study Limitations and Future Research

One limitation of the present study is its use of a self-report diagnostic measure (the PHQ) as the gold standard against which to assess the diagnostic utility of the BDI-II. Although the PHQ has been found to have very strong agreement with independent clinician-rated diagnoses (Spitzer et al., 1999), two such scores that are obtained by the same method (in this case, self-report) almost always have some degree of correlation. Therefore, the relationships between BDI scores and MDD diagnoses reported by the present study may be somewhat inflated due to shared method variance. Future studies should address this issue by using a structured interview, such as the Structured Clinical Interview for *DSM-IV* (SCID-IV; Spitzer, Williams, Gibbon, & First, 1994) or Schedule for Affective Disorders and Schizophrenia (SADS; Endicott & Spitzer, 1978).

Future research should also study the BDI-II's responsiveness to change in depressive symptomatology. Although the differences in BDI-II scores between people with and without MDD suggest that lower scores would follow treatment or spontaneous recovery, the cross-sectional data of the present study do not allow for a definitive conclusion. Therefore, longitudinal research is needed to study the validity of the BDI-II as a treatment outcome measure.

Some potential drawbacks to using the BDI-II should be mentioned. One is the cost of the BDI-II relative to some other measures of depression. Because the BDI-II must be purchased, it could be quite expensive to use for large-scale screening. Another potential drawback could arise if phone administration were desired, such as for treatment follow-up. The response format of the BDI-II requires respondents to choose among four statements, a format that may not fit well with phone interviews, although the specific phrases associated with each response option probably add to the information obtained.

Use of Somatic Items in a Medical Setting

A possible criticism of using the BDI-II in a medical setting is that the BDI-II includes somatic items, which may be confounded with physical illness rather than tapping into a depressive syndrome (Clark et al., 1983). In another study, a somatic-vegetative factor extracted from the BDI was correlated with illness severity and age, which was interpreted as support for excluding somatic items in medical settings (Karanci, 1988). An instrument has actually been developed for use in medical settings that is based on the BDI but excludes somatic items (see Beck, Steer, Ball, Ciervo, & Kabat, 1997). Further research has suggested that some somatic symptoms (e.g., diminished energy, appetite and sleep disturbance, and health worries) are accurate indices of depression, whereas other somatic symptoms (e.g., work inhibition, weight loss, loss of interest in sex) are not consistent with overall depression (Norris & Woehr, 1998). These latter items were deleted in the BDI revision.

Several findings in the present study challenge the argument that somatic items should be excluded from measures of depression used in medical settings. First, if scores from somatic items are less valid indicators of depression for a medical patient sample, then corrected item-total correlations should be lower for somatic items, which was not the case. Second, the predicted relationships between SF-20 subscales and BDI-II factor scores did not vary across factors including or not including somatic items. Third, all items from the BDI-II (including somatic items) made salient contributions to the second-order factor of depression. Therefore, regardless of whether somatic items on the BDI-II are sometimes confounded with physical illness, they still make salient contributions to a second-order factor of depression and are covarying with other cognitive and affective symptoms of depression.

Fourth, results from the ROC analysis indicated that the BDI-II total score, which includes somatic items, is quite efficacious for predicting the presence of MDD. Our finding of 92% overall accuracy in detecting MDD is strong evidence that the inclusion of somatic items in the total score did not detract from the predictive utility of the BDI-II scores. In fact, in comparison to the nonsomatic item BDI evaluated by Beck et al. (1997), the specificity was nearly identical, and the sensitivity was higher in the present study.

Finally, correlations between the SF-20 subscales and BDI-II factor scores revealed that the Somatic-Affective factor, as well as the Cognitive factor, correlated least with the SF-20 Physical Pain and Physical Functioning subscales, suggesting neither factor is confounded with physical illness.

These are five strong lines of evidence that somatic items should be included in measures of depression in medical settings.

Conclusion

The present study suggests that the BDI-II yields reliable and valid scores when used in a primary care setting. More extensive use of the BDI-II may improve detection, treatment, and understanding of depression in primary care medical patients.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Arnau, R. C. (1998, April). *Second-order factor analysis: Methods and interpretation.* Paper presented at the annual meeting of the Southwestern Psychological Association, New Orleans, LA. (ERIC Document Reproduction Service ED 417 216).

Beck, A. T., & Steer, R. A. (1993). *Manual for the Beck Depression Inventory.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., Ball, R., Ciervo, C. A., & Kabat, M. (1997). Use of the Beck Anxiety and Depression Inventories for primary care with medical outpatients. *Assessment, 4,* 211–219.

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

Buchwald, A. M., & Rudick-Davis, D. (1983). The symptoms of major depression. *Journal of Abnormal Psychology, 102,* 197–205.

Clark, D. C., Cavanaugh, S. V., & Gibbons, R. D. (1983). The core symptoms of depression in medical and psychiatric patients. *The Journal of Nervous and Mental Disease, 171,* 705–713.

Coyne, J. C., Fechner-Bates, S., & Schwenk, T. L. (1994). The prevalence, nature, and co-morbidity of depressive disorders in primary care. *General Hospital Psychiatry, 16,* 267–276.

Coyne, J. C., Fechner-Bates, S., & Schwenk, T. L. (1995). Nondetection of depression by primary care physicians reconsidered. *General Hospital Psychiatry, 16,* 267–276.

Crawford, C. B., & Koopman, P. (1973). A note on Horn's test for the number of factors in factor analysis. *Multivariate Behavioral Research, 8,* 117–125.

Depression Guideline Panel. (1993). *Depression in primary care* (Vol. 1: Detection and Diagnosis, Clinical Practice Guideline, No. 5). Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. (AHCPR Publication No. 93-0550).

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1988). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10,* 83–89.

Eisenberg, L. (1992). Treating depression and anxiety in primary care: Closing the gap between knowledge and practice. *New England Journal of Medicine, 326,* 1080–1084.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 35,* 837–844.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 29–36.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple solution. *British Journal of Statistical Psychology, 17,* 65–70.

Holmes, W. C. (1998). A short, psychiatric, case-finding measure for HIV seropositive outpatients. *Medical Care, 36,* 237–243.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

Karanci, N. A. (1988). Patterns of depression in medical patients and their relationship with causal attributions for illness. *Psychotherapy & Psychosomatics, 50,* 207–215.

Katon, W., Von Korff, M., Lin, E., Unutzer, J., Simon, G., Walker, E., Ludman, E., & Bush, T. (1997). Population-based care of depression: Effective disease management strategies to decrease prevalence. *General Hospital Psychiatry, 19,* 169–178.

Koenig, H. G., Cohen, H. J., Blazer, D. G., Krishnan, K. R. R., & Sibert, T. E. (1993). Profile of depressive symptoms in younger and older medical inpatients with major depression. *Journal of the American Geriatrics Society, 41,* 1169–1176.

Montono, C. B. (1994). Recognition and treatment of depression in a primary care setting. *Journal of Clinical Psychiatry, 55,* 18–34.

Muñoz, R. F., Hollon, S. D., McGrath, E., Rehm, L. P., & VandenBos, G. R. (1994). On the AHCPR Depression in Primary Care guidelines: Further considerations for practitioners. *American Psychologist, 49,* 42–61.

Norris, M. P., Meagher, M. W., & Arnau, R. C. (1999, August). *Structure of depression in older primary care patients.* Poster session presented at the 107th Annual Convention of the American Psychological Association, Boston, MA.

Norris, M. P., Snow-Turek, A. L., & Blankenship, L. (1995). Somatic depressive symptoms in the elderly: Contribution or confound? *Journal of Clinical Geropsychology, 1,* 5–17.

Norris, M. P., & Woehr, M. D. (1998). Differentiating somatic symptoms in the elderly: Depression, aging, or illness? *Journal of Mental Health and Aging, 4,* 347–359.

Osman, A., Downs, W. R., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory-II. *Journal of Psychopathology & Behavioral Assessment, 19,* 359–376.

Provalis Research. (1996). Simstat for Windows [Computer Software]. Montreal, Quebec, Canada: Author.

Regier, D. A., Narrow, W. E., Rae, D. S., Manderscheid, R. W., Locke, B. Z., & Goodwin, F. K. (1993). The de facto U.S. Mental and Addictive Disorders Service System: Epidemiological Catchment Area prospective 1-year prevalence rates of disorders and services. *Archives of General Psychiatry, 50,* 85–94.

Rey, J. M., Morris-Yates, A., & Stanislaw, H. (1992). Measuring the accuracy of diagnostic tests using receiver operating characteristics (ROC) analysis. *International Journal of Methods in Psychiatric Research, 2,* 39–50.

Romano, J. M., & Turner, J. A. (1985). Chronic pain and depression: Does the evidence support a relationship? *Psychological Bulletin, 97,* 18–34.

SAS Institute. (1990). SAS/IML software: Usage and reference (Version 6, First ed.). Cary, NC: Author.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

Spitzer, R. L., Kroenke, K., & Williams, J. B. W. (1999). Validation and utility of a self-report version of the PRIME-MD: The PHQ primary care study. *JAMA, 282,* 1737–1744.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1994). *Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition.* New York: New York State Psychiatric Institute.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology, 55,* 117–128.

Steer, R. A., Geetha, K., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *Journal of Psychopathology and Behavioral Assessment, 20,* 127–137.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Stewart, A. L., Hayes, R. D., & Ware, J. E. (1988). The MOS short-form general health survey: Reliability and validity in a patient population. *Medical Care, 26,* 724–735.

Swets, J. E. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285–1293.

Thompson, B. (1990). SECONDOR: A program that computes a second-order principal components analysis and various interpretation aids. *Educational and Psychological Measurement, 50,* 575–580.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57,* 5–19.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement, 45,* 203–209.

Thompson, B., & Borrello, G. M. (1992). Measuring second-order factors using confirmatory methods: An illustration with the Hendrick–Hendrick Love Instrument. *Educational and Psychological Measurement, 52,* 69–77.

Vasquez-Barquero, J. L., Herran, A., & Artal, J. (1997). Epidemiological research in primary care. *European Journal of Psychiatry, 11,* 133–143.

Wells, K. B., Stewart, A., Hays, R. D., Burnam, A., Rogers, W., Daniels, M., Berry, S., Greenfield, S., & Ware, J. (1989). The functioning and well-being of depressed patients. *Journal of the American Medical Association, 262,* 914–919.

Williamson, G. M., & Schulz, R. (1992). Pain, activity restriction, and symptoms of depression among community-residing elderly adults. *Journals of Gerontology: Psychological Sciences, 47,* 367–372.

Zung, W. W. (1990). The role of ratings scales in the identification and management of the depressed patient in the primary care setting. *Journal of Clinical Psychiatry, 51,* 72–76.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.