Review of the Beck Depression Inventory-II by PAUL A. ARBISI, Minneapolis VA Medical Center, Assistant Professor Department of Psychiatry and Assistant Clinical Professor Department of Psychology, University of Minnesota, Minneapolis, MN:

After over 35 years of nearly universal use, the Beck Depression Inventory (BDI) has undergone a major revision. The revised version of the Beck, the BDI-II, represents a significant improvement over the original instrument across all aspects of the instrument including content, psychometric validity, and external validity. The BDI was an effective measure of depressed mood that repeatedly demonstrated utility as evidenced by its widespread use in the clinic as well as by the frequent use of the BDI as a dependent measure in outcome studies of psychotherapy and antidepressant treatment (Piotrowski & Keller, 1989; Piotrowski & Lubin, 1990). The BDI-II should supplant the BDI and readily gain acceptance by surpassing its predecessor in use.

Despite the demonstrated utility of the Beck, times had changed and the diagnostic context within which the instrument was developed had altered considerably over the years (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Further, psychometrically, the BDI had some problems with certain items failing to discriminate adequately across the range of depression and other items showing gender bias (Santor, Ramsay, & Zuroff, 1994). Hence the time had come for a conceptual reassessment and psychometrically informed revision of the instrument. Indeed, a mid-course correction had occurred in 1987 as evidenced by the BDI-IA, a version that included rewording of 15 out of the 21 items (Beck & Steer, 1987). This version did not address the limited scope of depressive symptoms of the BDI nor the failure of the BDI to adhere to contemporary diagnostic criteria for depression as codified in the DSM-III. Further, consumers appeared to vote with their feet because, since the publication of the BDI-IA, the original Beck had been cited far more frequently in the literature than the BDI-IA. Therefore, the time had arrived for a major overhaul of the classic BDI and a retooling of the content to reflect diagnostic sensibilities of the 1990s.

In the main, the BDI-II accomplishes these goals and represents a highly successful revamping of a reliable standard. The BDI-II retains the 21-item format with four options under each item, ranging from not present (0) to severe (3). Relative to the BDI-IA, all but three items were altered in some way on the BDI-II. Items dropped from the BDI include body image change, work difficulty, weight loss, and somatic preoccupation. To replace the four lost items, the BDI-II includes the following new items: agitation, worthlessness, loss of energy, and concentration difficulty. The current item content includes: (a) sadness, (b) pessimism, (c) past failure, (d) loss of pleasure, (e) guilty feelings, (f) punishment feelings, (g) self-dislike, (h) self-criticalness, (i) suicidal thoughts or wishes, (j) crying, (k) agitation, (l) loss of interest, (m) indecisiveness, (n) worthlessness, (o) loss of energy, (p) changes in sleeping pattern, (q) irritability, (r) changes in appetite, (s) concentration difficulty, (t) tiredness or fatigue, and (u) loss of interest in sex. To further reflect DSM-IV diagnostic criteria for depression, both increases and decreases in appetite are assessed in the same item and both hypersomnia and hyposomnia are assessed in another item. And rather than the 1-week time period rated on the BDI, the BDI-II, consistent with DSM-IV, asks for ratings over the past 2 weeks.

The BDI-II retains the advantage of the BDI in its ease of administration (5-10 minutes) and the rather straightforward interpretive guidelines presented in the manual. At the same time, the advantage of a self-report instrument such as the BDI-II may also be a disadvantage. That is, there are no validity indicators contained on the BDI or the BDI-II and the ease of administration of a self-report lends itself to the deliberate tailoring of self-report and distortion of the results. Those of us engaged in clinical practice are often faced with clients who alter their presentation to forward a personal agenda that may not be shared with the clinician. The manual obliquely mentions this problem in an ambivalent and somewhat avoidant fashion. Under the heading, 'Memory and Response Sets,' the manual blithely discounts the potential problem of a distorted response set by attributing extreme elevation on the BDI-II to 'extreme negative thinking' which 'may be a central cognitive symptom of severe depression rather than a response set per se because patients with milder depression should show variation in their response ratings' (manual, p. 9). On the other hand, later in the manual, we are told that, 'In evaluating BDI-II scores, practitioners should keep in mind that all self-report inventories are subject to response bias' (p. 12). The latter is sound advice and should be highlighted under the heading of response bias.

The manual is well written and provides the reader with significant information regarding norms, factor structure, and notably, nonparametric item-option characteristic curves for each item. Indeed the latter inclusion incorporates the latest in item response theory, which appears to have guided the retention and deletion of items from the BDI (Santor et al., 1994).

Generally the psychometric properties of the BDI-II are quite sound. Coefficient alpha estimates of reliability for the BDI-II with outpatients was .92 and was .93 for the nonclinical sample. Corrected item-total correlation for the outpatient sample ranged from .39 (loss of interest in sex) to .70 (loss of pleasure), for the nonclinical college sample the lowest item-total correlation was .27 (loss of interest in sex) and the highest (.74 (self-dislike). The test-retest reliability coefficient across the period of a week was quite high at .93. The inclusion in the manual of item-option characteristic curves for each BDI-II item is of noted significance. Examination of these curves reveals that, for the most part, the ordinal position of the item options is appropriately assigned for 17 of the 21 items. However, the items addressing punishment feelings, suicidal thought or wishes, agitation, and loss of interest in sex did not display the anticipated rank order indicating ordinal increase in severity of depression across item options. Additionally, although improved over the BDI, Item 10 (crying) Option 3 does not clearly express a more severe level of depression than Option 2 (see Santor et al., 1994). Over all, however, the option choices within each item appear to function as intended across the severity dimension of depression.

The suggested guidelines and cut scores for the interpretation of the BDI-II and placement of individual scores into a range of depression severity are purported to have good sensitivity and moderate specificity, but test parameters such as positive and negative predictive power are not reported (i.e., given score X on the BDI-II, what is the probability that the individual meets criteria for a Major Depressive Disorder, of moderate severity?). According to the manual, the BDI-II was developed as a screening instrument for major depression and, accordingly, cut scores were derived through the use of receiver operating characteristic curves to maximize sensitivity. Of the 127 outpatients used to derive the cut scores, 57 met criteria for either single-episode or recurrent major depression. The relatively high base rate (45%) for major depression is a bit unrealistic for nonpsychiatric settings and will likely serve to inflate the test parameters. Cross validation of the cut scores on different samples with lower base rates of major depression is warranted due to the fact that a different base rate of major depression may result in a significant change in the proportion of correct decisions based on the suggested cut score (Meehl & Rosen, 1955). Consequently, until the suggested cut scores are cross validated in those populations, caution should be exercised when using the BDI-II as a screen in nonpsychiatric populations where the base rate for major depression may be substantially lower.

Concurrent validity evidence appears solid with the BDI-II demonstrating a moderately high correlation with the Hamilton Psychiatric Rating Scale for Depression-Revised (r = .71) in psychiatric outpatients. Of importance to the discriminative validity of the instrument was the relatively moderate correlation between the BDI-II and the Hamilton Rating Scale for Anxiety-Revised (r = .47). The manual reports mean BDI-II scores for various groups of psychiatric outpatients by diagnosis. As expected, outpatients had higher scores than college students. Further, individuals with mood disorders had higher scores than those individuals diagnosed with anxiety and adjustment disorders.

The BDI-II is a stronger instrument than the BDI with respect to its factor structure. A two-factor (Somatic-Affective and Cognitive) solution accounted for the majority of the common variance in both an outpatient psychiatric sample and a much smaller nonclinical college sample. Factor Analysis of the BDI-II in a larger nonclinical sample of college students resulted in Cognitive-Affective and Somatic-Vegetative main factors essentially replicating the findings presented in the manual and providing strong evidence for the overall stability of the factor structure across samples (Dozois, Dobson, & Ahnberg, 1998). Unfortunately several of the items such as sadness and crying shifted factor loadings depending upon the type of sample (clinical vs. nonclinical).

SUMMARY. The BDI-II represents a highly successful revision of an acknowledged standard in the measurement of depressed mood. The revision has improved upon the original by updating the items to reflect contemporary diagnostic criteria for depression and utilizing state-of-the-art psychometric techniques to improve the discriminative properties of the instrument. This degree of improvement is no small feat and the BDI-II deserves to replace the BDI as the single most widely used clinically administered instrument for the assessment of depression.

REVIEWER'S REFERENCES

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 52, 194-216.

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.

Beck, A. T., & Steer, R. A. (1987). Beck Depression Inventory manual. San Antonio, TX: The Psychological Corporation.

Piotrowski, C., & Keller, J. W. (1989). Psychological testing in outpatient mental health facilities: A national study. Professional Psychology: Research and Practice, 20, 423-425.

Piotrowski, C., & Lubin, B. (1990). Assessment practices of health psychologists; Survey of APA Division 38 clinicians. Professional Psychology: Research and Practice, 21, 99-106.

Santor, D. A., Ramsay, J. O., & Zuroff, D. C. (1994). Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights. Psychological Assessment, 6, 255-270.

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.