To perform a systematic review of the utility of the Beck Depression Inventory for detecting depression in medical settings, this article focuses on the revised version of the scale (Beck Depression Inventory-II), which was reformulated according to the DSM-IV criteria for major depression. We examined relevant investigations with the Beck Depression Inventory-II for measuring depression in medical settings to provide guidelines for practicing clinicians. Considering the inclusion and exclusion criteria seventy articles were retained. Validation studies of the Beck Depression Inventory-II, in both primary care and hospital settings, were found for clinics of cardiology, neurology, obstetrics, brain injury, nephrology, chronic pain, chronic fatigue, oncology, and infectious disease. The Beck Depression Inventory-II showed high reliability and good correlation with measures of depression and anxiety. Its threshold for detecting depression varied according to the type of patients, suggesting the need for adjusted cut-off points. The somatic and cognitive-affective dimension described the latent structure of the instrument. The Beck Depression Inventory-II can be easily adapted in most clinical conditions for detecting major depression and recommending an appropriate intervention. Although this scale represents a sound path for detecting depression in patients with medical conditions, the clinician should seek evidence for how to interpret the score before using the Beck Depression Inventory-II to make clinical decisions.

KEYWORDS: Beck Depression Inventory; Depression; Medical Illness; Psychometric Scale; Screening; Validation Study.

INTRODUCTION

Patients with chronic medical illness have a high prevalence of major depressive illness (1). Depressive symptoms may co-occur with serious medical illnesses, such as heart disease, stroke, cancer, neurological disease, HIV infection, and diabetes (1-3). The functional impairment associated with medical illnesses often causes depression. Patients who present depression along with medical illness tend to have more severe symptoms, more difficulty adjusting to their health condition, and more medical costs than patients who do not have co-existing depression (2). While prompt treatment of depression can improve the outcome of the co-occurring physical illness, proper and early recognition of treatable depression can result in a faster recovery and can shorten the patient’s hospital stay. Formal assessment of depression by a liaison psychiatrist or clinician-administered instruments, such as the Hamilton Depression Rating Scale (4) and the Montgomery-Åsberg Depression Rating Scale (5), are onerous to implement in routine clinical settings. In contrast, self-report measures for depression can be cost-effective for use in busy specialty medical clinics. Throughout the second half of the 20th century, along with the discovery of effective antidepressant drugs and the development of cognitive-behavioral therapy, several patient-rated assessment scales for detecting depression were proposed. Popular instruments include the Beck Depression Inventory (BDI) (6), the Self-Rating Depression Scale (7), the Center for Epidemiologic Studies Depression Scale (8), the Patient Health Questionnaire-9 (9), the Inventory of Depressive Symptomatology (10), and the Depression in the Medically Ill (11). Alternative scales have been developed to measure depression in specific populations, such as postpartum women (12) and patients with schizophrenia (13). Other scales have been devoted to quantify depression in specific age groups, such as adolescents (14) and the elderly (15). The utility of these
scales in the medically ill is challenging because the frequent presence of somatic symptoms in physical diseases can mislead their score interpretation. If the clinician is unable to decide which existing instrument to use and how to interpret the results, the advancement of self-rating scales can represent a step backward.

Among the investigations on using self-assessment measures to evaluate depression, the BDI outnumbered the other measures in the amount of published research: there are more than 7,000 studies so far using this scale. Aaron T. Beck and colleagues developed the 21-item BDI in 1961 to aid clinicians in the assessment of psychotherapy for depression (6). The easy applicability and psychometric soundness of this scale have popularized its use in a variety of samples (16-19) and in healthcare settings worldwide (20-22). This inventory has received two major revisions: in 1978 as BDI-I (23) and in 1996 as BDI-II (24). This later reformulation covers psychological and somatic manifestations of a two-week major depressive episode, as operationalized in the DSM-IV (25). Four items of the BDI-I (weight loss, distorted body image, somatic preoccupation, and inability to work) were replaced with agitation, worthlessness, difficulty concentrating, and energy loss to assess the intensity of depression. The items of appetite and sleep changes were amended to evaluate the increase and decrease in depression-related vegetative behaviors (24,26-28). Different from the original version, which intended to measure negative cognitions of depression, the BDI-II does not reflect any particular theory of depression. The English version of BDI-II has been translated and validated in 17 languages so far, and it is used among countries in Europe, the Middle East, Asia, and Latin America (29-32).

Investigations on depression and its instrumentation must be considered in view of the pressure for evidence-based decisions in clinical practice and the information explosion of the literature. Recently, the BDI-II has been ever-increasingly used in the medically ill to evaluate depressive states that occur at high prevalence in healthcare settings. The authors systematically reviewed the validity of the BDI-II to quantify the severity of depression among medical patients and discuss the interpretation of its metric conventions. The performance of the BDI-II (and its short version BDI-FS) among patients with medical illnesses who often present somatic complaints is contrasted with its performance among non-medical patients, among whom psychological symptoms are the most prominent features.

## METHODS

Both investigators, with previous experience on psychometric instruments, conducted this systematic review by searching the Web of Sciences (ISI), Medline, and PsycINFO databases. The following MeSH terms were used to scan studies through the search builder of each database: “valid*” OR “reliab*” OR “sensitiv*” OR “specific*” OR “concurrent” OR “divergent” OR “convergent” OR “factor analysis”. Following the search, we filtered articles containing the term “Beck Depression Inventory” published during the time period “1/1/1996 to 10/10/2012”. There was no language or age range restriction. The initial search resulted in 822 retrieved articles, with 409 from ISI, 328 from Medline, and 85 from PsycINFO. The reference sections of the review articles of the depression instruments (33-35) and book chapters (36-38) were examined to identify potential studies. Additional efforts to locate relevant studies by hand and to contact experts in the field identified seven psychometric articles on medical samples, totaling 829 articles.

After checking for duplication and overlap, 528 articles remained in the list. Filtering non-medical articles, we eliminated 170 articles in which “student,” “psychiatric,” or “community” was mentioned in the title or abstract. The retained 358 articles were screened for eligibility by reading the abstract. Two articles were not accessible, even upon request to the author, resulting in 356 full-text articles that were assessed for eligibility.

The exclusion criteria were as follows: (1) non-psychometric studies, such as clinical trials, editorials, letters, reviews, meta-analyses, practice guideline, randomized controlled trials, and case reports; (2) non-medical samples (student, psychiatric, or non-clinical); (3) small sample size (N<30); (4) BDI-I; and (5) reanalysis or duplicated analysis of an original dataset. The sample was considered “non-clinical” when study participants consisted of workers, caregivers, and community dwellers. Regardless of the nosological controversy of chronic fatigue syndrome and chronic pain as medical illnesses, these conditions were included due to their high occurrence in healthcare settings. Samples with less than 30 participants were only retained when the study addressed a very important problem, such as between-version comparison or content analysis. A summary analysis of the complete sample was preferable when multiple analyses were available (such as separate reports by gender, ethnicity, or depressed versus non-depressed groups).

The reasons for excluding 286 articles were as following: 147 studies did not contain the original data using the BDI-II (167 non-psychometric studies and seven reviews); 95 studies utilized non-medical samples (34 student samples, 31 psychiatric samples, and 30 non-clinical samples); 13 studies provided a reanalysis or secondary data analysis; three studies used BDI-I; and one study had a small sample size. The final list resulted in 70 articles that are dedicated to investigating the psychometric performance of the BDI-II in medical patients. The flowchart in Figure 1 displays each step of the search process.

Studies on medical diseases were grouped according to the sample recruitment source as outpatients or primary care (k = 52) and hospital (k = 12) (Table 1). Studies investigating the short version BDI-FS (k = 10) are displayed separately. Four studies reported data on both BDI-II and BDI-FS. Several investigations did not provide a clear description of the healthcare setting or recruited participants from different levels of health service. Likewise, the heterogeneous selection of patients might reflect different groups of participants or stages of disease course. Sixteen studies reported a sample size with less than 100 respondents, but all of the studies had more than the minimum of 30 subjects.

Among the 70 retained studies, the BDI-II was administered to adults in primary care (k = 4) and clinics of cardiology (k = 12), neurology (k = 12), obstetrics (k = 8), brain injury (k = 6), nephrology (k = 5), chronic pain (k = 4), chronic fatigue (k = 4), oncology (k = 5), and infectious disease (k = 3). Only two studies assessed adolescent medical patients (39,40).

Almost all of the identified studies were published after 2000, and the great majority (approximately 64%) of studies...
was published in the past five years, suggesting a recent trend for using the BDI-II in medical settings. Nearly 70% of the articles applied the English version of BDI-II, but 13 non-English versions of the scale were found.

**Overview**

The BDI-II performed well in adult patients with a wide array of medical diseases (Table 1). For the purpose of comparison, data from Beck’s studies on non-medical and medical samples (24,26) are listed as normative references. Usually, non-patient samples reported the item scores in the lower part of the range of possible scores (from 0 to 3), with a skewed distribution of item scores. Based on scores of 500 psychiatric outpatients, Beck et al. (24) suggested the following ranges of BDI-II cut-off scores for depression: 0–13 (minimal), 14–19 (mild), 20–28 (moderate), and 29–63 (severe). As an example, the mean score of the BDI-II in samples with mood disorder was $M = 26.6$, and the mean scores for major depressive episode, recurrent depression, and dysthymia were 28.1, 29.4, and 24.0, respectively.

Confirming the expectation that medical patients would report more somatic symptoms, most of the investigations reported a slightly higher mean total score for medical patients than non-patients (Table 1), but scores were still around or below the threshold of 13/14 that is recommended by Beck to detect mild depression. Exceptions of this observation were studies on chronic pain (29,61,70,77), with mean total scores ranging from 17.2 to 26.9. The type of respondents might influence item endorsement and the scale total score.

In comparison with the previous version, the item characteristics of the BDI-II have been changed in terms of endorsement rate, homogeneity, and content coverage (34). The homogeneity of the scale was described for 17 of 21 items in the original study (24), showing acceptable item-total
| Authors, year | Language | N | Sample description | %W | Mean Score (SD) | Alpha |
|--------------|----------|---|-------------------|----|----------------|-------|
| Beck et al., 1996 (24) | English | 120 | College students | 44 | 12.6 (9.9) | 0.93 |
| | 500 | Psychiatric outpatients | 62 | 22.5 (12.8) | 0.92 |
| Arnarson et al., 2008 (41) | Icelandic | 248 | Adult outpatients | 82 | 21.3 (12.2) | 0.93 |
| Arnau et al., 2001 (42) | English | 333 | Adult - primary care | 69 | 8.7 (9.4) | 0.94 |
| Brown et al., 2012 (43) | English | 111 | Chronic fatigue outpatients | 83 | 17.7 (9.1) | 0.89 |
| Beck & Gable, 2001 (44) | English | 150 | Postpartum outpatients | 100 | NR | 0.91 |
| Bunevicius et al., 2012 (45) | Lithuanian | 522 | Coronary outpatients | 28 | 11.0 (8.2) | 0.85 |
| Carney et al., 2009 (46) | English | 140 | Insomnia outpatients | 74 | 14.1 (10.2) | 0.91 |
| Carvalho Bos et al., 2009 (47) | Portuguese | 331 | Pregnancy outpatients | 100 | NR | 0.88 |
| Chaudron et al., 2010 (48) | English | 198 | Postpartum outpatients | 100 | NR | NR |
| Chilcott et al., 2008 (49) | English | 40 | Renal hemodialysis outpatients | 40 | 11.1-12.9 (9.3-9.4) | NR |
| Chilcott et al., 2011 (50) | English | 460 | Renal disease outpatients | 35 | 11.9 (8.3) | NR |
| Chung et al., 2010 (51) | Chinese | 62 | Heart disease outpatients | 31 | 18.2 (7.9) | NR |
| Corbi-Soto et al., 2011 (52) | French | 206 | Chronic pain outpatients | 53 | 17.2 (11.5) | 0.84 |
| Dbouk et al., 2008 (53) | English | 129 | Hepatitis C outpatients | 50 | 17.1 (11.6) | NR |
| de Souza et al., 2010 (54) | Spanish | 50 | Huntington’s disease | 48 | 8.8 (8.9) M 26.8 (6.9) D | NR |
| del Pino Pérez et al., 2012 (55) | Spanish | 205 | Coronary outpatients | 26 | 9.2 (7.6) | NR |
| Dutton et al., 2004; | English | 220 | Adult - primary care | 52 | 12.6 (10.4) | 0.90 |
| Grothe et al., 2005 (56) | English | 98 | Traumatic brain injury (mild) | 55 | 12.2 (9.6) | NR |
| | 228 | Traumatic brain injury (moderate to severe) | 33 | 9.7 (8.1) | NR |
| Frause-Smith & Lespérance, 2008 (57) | English/French | 804 | Coronary heart disease outpatients | 34 | 9.4 (8.9) M 17.8 (8.7) D | NR |
| Griffith et al., 2005 (58) | English | 132 | Epilepsy outpatients | 72 | 15.9 (11.1) | NR |
| Hamid et al., 2004 (59) | Arabic | 493 | Women - primary care | 100 | 13.0 (8.1) | NR |
| Harris & D’Eon, 2008 (60) | English | 481 | Chronic pain outpatients | 58 | 26.9 (11.7) | 0.82 |
| Hayden et al., 2012 (61) | English | 83 | Obese bariatric outpatients | 71 | 13.4 (9.1) | 0.89 |
| Jones et al., 2005 (62) | English | 174 | Epilepsy outpatients | 66 | NR | 0.94 |
| Kanner et al., 2010 (63) | English | 193 | Epilepsy outpatients | 68 | 10.6 (6.3) | NR |
| King et al., 2012 (64) | English | 489 | Traumatic brain injury | 10 | 19.7 (11.8) | NR |
| Kiropoulos et al., 2012 (65) | English | 152 | Coronary heart disease outpatients | 34 | 9.4 (8.9) M 17.8 (8.7) D | NR |
| Kirsch-Darrow et al., 2011 (66) | English | 161 | Parkinson outpatients | 31 | 9.5 (7.2) | 0.89 |
| Ko et al., 2012 (67) | Korean | 121 | Epilepsy outpatients | 35 | 9.7 (6.3) M 29.9 (11.7) D | NR |
| Lipps et al., 2010 (68) | English | 191 | HIV infection outpatients | 61 | 14.1 (11.0) M 10.2 (9.1) M | 0.89 |
| Lopez et al., 2012 (69) | English | 345 | Chronic pain outpatients | 0 | 23.0 (12.2) | 0.93 |
| Masuda et al., 2012 (70) | Japanese | 327 | Myasthenia gravis outpatients | 67 | 11.3 (7.9) | NR |
| Neitzer et al., 2012 (71) | English | 150 | Renal hemodialysis outpatients | 48 | 12.3 (10.8) | NR |
| Ooms et al., 2011 (72) | Dutch | 136 | Tinnitus outpatients | 35 | 11.3 (9.5) | NR |
| Osa da et al., 2011 (73) | Japanese | 56 | Fibromyalgia outpatients | 86 | NR | NR |
| Patterson et al., 2011 (74) | English | 671 | Hepatitis C outpatients | 3 | 16.2 (12.2) | 0.84-0.91 |
| Penley et al., 2003 (75) | English/Spanish | 122 | Chronic renal outpatients | 41 | 15.0 (12.5) | 0.92 |
| Pereira et al., 2011 (76) | Portuguese | 503 | Pregnant outpatients | 100 | NR | NR |
| Poole et al., 2009 (77) | English | 1227 | Chronic pain outpatients | 62 | 24.7 (11.6) | 0.92 |
| Rampling et al., 2012 (78) | English | 266 | Epilepsy outpatients | 59 | NR | 0.94 |
| Roe buck-Spencer, 2006 (79) | English | 60 | Systemic lupus erythematosus outpatients | 80 | NR | NR |
| Su et al., 2007 (80) | Chinese | 185 | Pregnant outpatients | 100 | 7.0 (5.0) M 17.0 (10.2) D | NR |
| Suzuki et al., 2011 (81) | Japanese | 287 | Myasthenia gravis outpatients | 67 | 11.1 (8.1) | NR |
| Tandon et al., 2012 (82) | English | 95 | Perinatal women | 100 | NR | 0.9 |
| Teng et al., 2005 (83) | Chinese | 203 | Postpartum outpatients | 100 | 7.8 (6.3) M 25.8 (10.4) D | NR |
| Turner et al., 2012 (84) | English | 72 | Stroke outpatients | 47 | 13.4 (12.9) | 0.94 |
| Turner-Stokes et al., 2005 (85) | English | 114 | Brain injury outpatients | 43 | 17.2 (11.5) | NR |
| Viljoen et al., 2003 (86) | English | 127 | Adult - primary care | 63 | NR | NR |
| Wan Mahmud et al., 2004 (87) | Malay | 61 | Postpartum I outpatients | 100 | 4.4 (5.5) | 0.89 |
| War menhoven et al., 2012 (88) | Dutch | 46 | Cancer outpatients | 43 | 14.7 (9.9) | NR |
| Williams et al., 2012 (89) | English | 229 | Parkinson disease outpatients | 33 | 6.5 (5.2) M 14.7 (7.4) D | 0.90 |
| Young et al., 2007 (90) | English | 194 | Cardiac outpatients | 35 | 8.6-13.4 (7.7-12.3) | NR |
| Zahodne et al., 2009 (91) | English | 71 | Parkinson disease outpatients | 32 | 11.7 (7.9) | NR |
| Di Benedetto et al., 2006 (92) | English | 81 | Acute cardiac syndrome | 19 | NR | > 0.90 |
| Gorenstein et al., 2011 (93) | Portuguese | 334 | Adult - hospitalized | 48 | 12.2 (11.6) | 0.91 |
| | | | 170 physically disabled | 14.5 (11.2) | NR |
| | | | 164 intellectually disabled | 9.7 (11.4) | NR |

Table 1 - Description of psychometric studies of the Beck Depression Inventory-II in medical samples by language version, sample size (N), sample description, gender distribution (%W), mean score (SD), and reliability (Cronbach’s alpha).
Table 1 - Continued.

| Authors, year | Language | N | Sample description | %W | Mean Score (SD) | Alpha |
|---------------|----------|---|--------------------|----|----------------|-------|
| Homifar et al., 2009 (94) | English | 52 | Traumatic brain injury * | 10 | 25 (14.6) | NR |
| Huffman et al., 2010 (95) | English | 131 | Myocardial infarction | 20 | 9.8 (9.4) | NR |
| Jamroz-Wisniewska et al., 2007 (96) | Polish | 104 | Multiple sclerosis | 74 | 14.4 (9.2) | NR |
| Low & Hubley, 2007 (97) | English | 119 | Coronary disease | 25 | 8.0 (7.1) | 0.89 |
| Pietsch et al., 2012 (40) | German | 314 | Adolescents patients* | 60 | 7.5 (6.5) | 0.91 |
| | | | (252 hospital inpatients) | | | |
| Rowland et al., 2005 (98) | English | 51 | Traumatic brain injury | 28 | 5.6 ND 20.1 D | NR |
| Siegert et al., 2009 (99) | English | 353 | Neurological diseases | 40 | 13.6 (10.1) | 0.89 |
| Thomas et al., 2008 (100) | English | 50 | Stroke | 38 | 12.7 (8.9) | NR |
| Thomas et al., 2008 (101) | English/French | 477 | Acute myocardial infarction | 17 | 9.2 (7.9) | NR |
| Tully et al., 2011 (102) | English | 226 | Cardiac heart disease | 17 | 8.6 (6.2) a | 0.85 |
| | | | | | 9.1 (6.4) b | 0.87 |

BDI Fast Screen version (k = 10)

| Authors, year | Language | N | Sample description | %W | Mean Score (SD) | Alpha |
|---------------|----------|---|--------------------|----|----------------|-------|
| Beck et al., 1997 (26) | English | 50 | Medical inpatients | 60 | 5.8 (4.5) | 0.86 |
| Brown et al., 2012 (43) | English | 111 | Chronic fatigue outpatients | 83 | 4.3 (3.2) | NR |
| Neitzer et al., 2012 (72) | English | 146 | Renal hemodialysis outpatients | 48 | 2.7 (3.4) | NR |
| Pietsch et al., 2012 (40) | German | 314 | Adolescents* | 60 | 1.9 (2.4) ND 8.1 (3.5) D | 0.82 |
| | | | (252 hospital inpatients) | | | |
| Poole et al., 2009 (103) | English | 1227 | Chronic pain outpatients | 62 | 7.1 (4.30) | 0.84 |
| Scheinthal et al., 2001 (104) | English | 75 | Geriatric outpatients | 56 | 2.3 (3.1) | 0.83 |
| Servaes et al., 2000 (105) | Dutch | 85 | Disease-free cancer outpatients | 43.5 | 0.4-2.3 (0.9-1.8) | NR |
| | | | Chronic fatigue outpatients | 50 | 2.6 (1.8) | |
| Servaes et al., 2002 (106) | Dutch | 57 | Disease-free breast cancer outpatients | 100 | 2.3-4.2 (2.2-3.9) | NR |
| | | | Chronic fatigue outpatients | 100 | 3.3 (2.6) | |
| Steer et al., 1999 (107) | English | 120 | Medical outpatients | 50 | 2.2 (3.0) | 0.85 |
| Winter et al., 1999 (39) | English | 100 | Adolescent outpatients | 50 | 1.9 (3.1) | 0.88 |

N: sample size; %W: percentage of women; SD: standard deviation; Alpha: Cronbach’s alpha coefficient of internal consistency; NR: not reported.
M: men; W: women; ND: non-depressed; D: depressed; *: pre-surgery; b: post-surgery.
†Mixed sample of in- and outpatients.
Separate analysis of the short version of the BDI-II in the same study.
IQR: interquartile range.

correlations of $r_{11} \geq 0.5$ (108). Different item endorsements and coverage are reported for different versions of the instrument: substantial item-total correlation was described for 15 items in the Brazilian-Portuguese version (93) and 10 items in the Arabic version (32). Direct comparison of the scores between different language versions should be avoided.

In contrast with patient samples, somatic items, such as “change in sleeping pattern” and “change in appetite,” presented low scores for non-clinical samples. However, “tiredness or fatigue,” might present special clinical significance in patients with chronic fatigue syndrome (43) or cardiac coronary disease (45,51). Regardless of the severity of depression, the item “loss of sexual interest” displayed the worst item-total correlation, although it was significantly related to the whole construct under consideration (23,24). Thombs et al. (101) suggested that the assessment of symptom severity with BDI–II would be substantially biased in medically ill patients due to the misattribution of symptom variance. The item “suicidal thoughts” was the least reported item among non-medical settings; however, a substantial correlation still demonstrates its contribution to depression (23,24). Investigations on the ability of separate items, e.g., “pessimism” and “loss of energy,” to predict disease outcome or treatment response can help clinicians in the management of depression. The contribution of self-rated somatic vs. cognitive symptoms in medical samples should be clarified by item analysis to identify whether items are appropriately assigned to a scale.

BDI-Fast Screen

Experts view somatic symptoms among medical patients as the harbinger of depression and anxiety in the healthcare setting (3,109-111). Preferably, the assessment of depression in patients with medical illness should avoid confounding physical symptoms. The correct identification of comorbid depressive disorders in medical patients is crucial in understanding its origin and in controlling the physical symptom burden. Two measures were designed with the objective of eliminating somatic items. The first proposed measure is the Hospital Anxiety Depression Scale (HADS) (112), which has a seven-item depression subscale. Despite the lack of comprehensive data on its psychometric properties (113) and challenges to its factorial validity (114), the HADS remained widely used as a research measure of depression in the medically ill.
The seven-item BDI for Primary Care (BDI-PC) (26) was developed in 1997 after removing somatic items, such as fatigue and sleep problems, from the BDI. This version was projected for evaluating depression in patients whose behavioral and somatic symptoms are attributable to biological, medical, alcohol, and/or substance abuse problems that may confound the diagnosis of depression. The BDI-PC was later renamed the BDI-F (Fast Screen for Medical Patients (BDI-FS)), and it consists of items 1 to 4 and 7 to 9 of the BDI-II (27).

The BDI-FS requires less than five minutes for completion, and scoring is similar to the BDI-II. For interpretation, the manual suggests that scores 0–3 indicate minimal depression; 4–6 indicate mild depression; 7–9 indicate moderate depression; and 10–21 indicate severe depression (27). Validation studies (k = 10) have demonstrated the ability of this non-somatic scale to discriminate depressed vs. non-depressed medical patients (39,26,104,107), chronic pain patients (103), and conditions where fatigue is a prominent feature (43,105,106). Less popular than its full version, more investigations are needed to establish the utility of this short version in medical settings before recommending its extensive use.

Reliability
Thirty-seven of 70 retrieved psychometric articles (52.9%) did not report reliability coefficients for the data. In comparison to the internal consistency of previous versions of the BDI (average Cronbach’s alpha coefficient of approximately 0.85) (23), the reliability of the BDI-II among medical samples was satisfactory, with an alpha of approximately 0.9, ranging between 0.84 and 0.94 (Table 1). In addition, Beck (26) reported a coefficient of 0.86 for the BDI-FS, and other studies reported the coefficient ranging from 0.82–0.88 (39,40).

No information on the retest reliability is available for medical samples. However, the stability of the BDI-II, as expressed by retest coefficients of Pearson’s r of 0.92 and 0.93, was reported by Beck and colleagues (24) for psychiatric and student samples, respectively. Further evidence of acceptable stability through re-application of the BDI-II was demonstrated for student samples (range: 0.73-0.96) (115,116).

The retest effect – that is, lower scores on the second application, even without intervention – may affect the reliability of BDI-II in healthcare settings. This effect could be unrelated to a true change in severity and could be purely the result of the measurement process. Although this fact would not preclude using this scale in follow-up or interventional studies among medical patients, nothing should be stated concerning the scale performance in this respect. Therefore, clinicians should be careful when making important treatment decisions based on non-empirical information assumed from non-clinical samples.

Item Response Theory
Most validation studies of BDI-II were analyzed in accordance with the classic test theory, assuming a true score for each respondent’s summed score and disregarding the measurement error. In other words, two individuals with the same total score may differ greatly in terms of relative severity and frequency of symptoms. This discrepancy might be particularly taxing in medical settings, where physical symptoms are common complaints and overlap with “true” depression-related somatic symptoms.

In the last decades, the item response theory (IRT) is an increasingly used method in psychometrics, in addition to the dominant classic test theory of true score paradigm. Briefly, the IRT distinguishes between moderate and severe cases of depression using item-level analysis to account for measurement error (117). The response of a respondent for a given ability should be modeled to each item in the test. For example, when a given depression scale is composed only of items that measure mild depression, this instrument would have great difficulty identifying severe depression because both levels of severity should be characterized by high scores on all items. In addition, if items assessing psychological and physical symptoms were only loosely related, a single score would not distinguish between two potentially different groups of depressed patients - with primarily psychological or with primarily vegetative symptoms. This scenario is particularly pressing in medical settings that are investigating clinical changes in depressive syndrome.

Seigert and colleagues (99) reported an illuminating study after examining each BDI-II item for differential item functioning in a neurological sample (n = 315). The authors identified misfits to model expectations for three items that seemed to measure different dimensions: changes in sleeping pattern; interest in meals; and loss of interest in sex. These vegetative items were removed and re-scored in an iterative fashion to the scale. In the real world, the likelihood of receiving a rating of 1 on the insomnia item was essentially the same, regardless of the overall severity of depression, but the likelihood of receiving a rating of 3 on sad mood could be low, even when overall depression was severe.

Waller and colleagues (118) investigated the latent structure of the BDI-II through differential item functioning and item level factor analysis in samples of women with breast cancer and women with clinical depression. Items of negative cognitions about the self, e.g., worthlessness, self-dislike, and punishment feelings, were less likely to be reported by breast cancer patients than depressed patients. Negative cognitions about the self appear to be related to different factors in breast cancer. The analyses also found many differences at both the item and factor scale levels, suggesting caution when interpreting the BDI-II in breast cancer patients.

These studies advocate that the rating scheme is not ideal for many BDI-II items, thus affecting the scale’s capacity to detect change in medical conditions. Systematic IRT analysis of the BDI-II items can strengthen the scale coverage in assessing heterogeneous depressive conditions among medical patients.

Convergent and Divergent Validity
Table 2 displays the studies that compared the BDI-II with scales measuring depression, anxiety, and miscellaneous constructs as criteria that were determined at essentially the same time to check for concurrent validity. The convergent validity between the BDI-II and the BDI-I was 0.93 (28). The shorter version, BDI-FS, also presented an acceptable correlation of 0.85 (72). In general, the overlap of the construct measured by BDI-II with other widely used scales to assess depression, e.g., the Center for Epidemiologic Studies of Depression, the Hamilton Depression Rating Scale, the Edinburg Postnatal Depression Scale, and the Hospital Anxiety and Depression Scale-Depression, was adequate and ranged from 0.62 to 0.81 (Table 2).
Additionally, the convergent validity between the BDI-II and scales that assess anxiety was significant and differed across comparison instruments: Beck Anxiety Inventory (0.60) (24,41), Hamilton’s Anxiety Rating Scale (0.47) (24), State-Trait Anxiety Inventory (0.83) (92), Penn State Worry Questionnaire (0.61) (41), and Hospital Anxiety and Depression Scale-Anxiety (0.65) (41). These results were expected due to the extent that anxiety symptoms were highly comorbid with depressive symptoms or that they could be attributed to the characteristics of the compared instruments. As a broad indicator of mental health, a high score on the BDI scale could also be explained by other disorders, physical illnesses, or social problems (69). Most likely, the construct covered by the BDI-II is beyond the “pure” depressive-type of psychopathology. As such, the convergent validity of the scale with hopelessness (24) and fatigue (105) was also substantial. In the medical setting, the clinician should not assume depression as a primary issue when BDI-II is used without a thorough clinical assessment.

Concerning divergent validity, studies have indicated poor correlation (r<0.4) with instruments assessing chronic pain (61), physical health (43), and substance use disorders (119). Suicidal ideation, which is one of core features of depression and an item on the BDI-II, was only poorly correlated with the instrument (24).

### Criterion-oriented Validity

Psychometric experts view the interpretation of the raw scores on tests, such as the BDI-II, as problematic, unless they are converted into standardized scores (e.g., T score or stanine method) (108,120). No known standardized norms have been reported for the BDI-II to date. As an alternative to the norm-referenced method, the criterion-referenced method is the most widespread practice for interpreting BDI-II scores. Normally, the total score is compared with a cut-off score established according to a gold-standard criterion (e.g., clinical assessment or structured interview).

When clinicians intend to screen probable cases of major depression in medical settings, the sensitivity should be

---

**Table 2 - Concurrent validity of the Beck Depression Inventory-II with measures of depression, anxiety, and other miscellaneous constructs in medical samples.**

| Concurrent instrument | r       | Study |
|-----------------------|---------|-------|
| **Depression measure**|         |       |
| BDI-I                 | 0.93    | 28    |
| BDI-FS                | 0.85    | 72    |
| HADS-D                | 0.62 - 0.71 | 261, 41 |
| CES-D                 | 0.72 - 0.87 | 29, 41, 52, 63, 69 |
| HRSD                  | 0.71 - 0.75 | 24, 87 |
| EPDS                  | 0.72 - 0.82 | 44, 83, 87 |
| GDS                   | 0.81    | 104†  |
| PHQ                   | 0.84    | 52    |
| CDS                   | 0.65; 0.69 | 66, 92 |
| POMS-D                | 0.77    | 59    |
| PDSS                  | 0.77    | 44, 76 |
| DISC                  | 0.66    | 85    |
| NGRS                  | 0.65    | 41    |
| **Anxiety measure**   |         |       |
| BAI                   | 0.60    | 24, 41 |
| HARS                  | 0.47    | 24    |
| STAI                  | 0.64; 0.83 | 66, 92 |
| PSWQ                  | 0.61    | 41    |
| HADS-D                | 0.65    | 41    |
| **Miscellaneous**     |         |       |
| SSI                   | 0.37    | 24    |
| BHS                   | 0.58    | 24    |
| MPP-O-PRI             | 0.52    | 61    |
| SF-36 MH              | 0.45 - 0.70 | 431, 57 |
| SF-36 PH              | 0.12 - 0.29 | 431, 57 |
| SPS                   | 0.39 - 0.42 | 69    |
| CIS-F                 | 0.58    | 105   |
| NDDI-E                | 0.81 - 0.85 | 64, 68 |
| NSI                   | 0.77    | 65    |
| MG-QOL                | 0.52    | 71    |
| JIQI                  | 0.58    | 74    |
| ANAM                  | 0.67    | 79    |
| SCOR                  | 0.54 - 0.80 | 100   |
| STOP-D                | 0.83    | 90    |
| LARS                  | 0.45    | 91    |
| AS                    | 0.58    | 91    |
| UPDRS-III             | 0.33    | 91    |

r: Pearson’s product moment correlation. Negative correlation is omitted in the numerical value.

†The concurrent validity refers to the BDI-FS version.

* A complete list of retrieved studies can be obtained from the authors upon request.
viewed as the most important indicator to minimize the chance of false-negative cases (Table 3). Sometimes, the BDI-II can overestimate the prevalence of depression in particular conditions, e.g., medically ill patients would record more items that address physical complaints. According to the samples, medical studies have reported good performance with high sensitivity (from 72% to 100%). Occasionally, the researcher might want to improve the specificity to select a pure sample of depressed patients. For research purposes, Beck et al. (24) recommended raising the cut-off score to 17 to obtain homogeneous samples of depressed individuals.

According to Table 3, the best cut-off to indicate cases of depressive syndrome in medical samples was established on the ground of the unique characteristics of the sample. The possible threshold ranged widely, from 7 to 22 (89,103).

For example, Poole et al. (103) found that raising the BDI-II cut-off score to 22 could reduce the number of false-positives produced by the uneven item response of chronic pain patients. Consequently, the researcher can change the flexibility of the cut-off score by comparing different thresholds for a new sample or study purpose.

A significant diagnostic accuracy of 82% and higher, as expressed by the area under the receiver operating characteristics (ROC) curve, was calculated according to the tradeoff between sensitivity and specificity. However, the ability of a scale to differentiate between depressive vs. non-depressive groups depends not only on the sensitivity and specificity of its cut-off scores but also on the frequency of the disorder in the samples that are being studied. In addition, sources of threshold variation may depend on the type of the sample (outpatient or hospitalized), medical

---

### Table 3 - Criterion validity and cut-off point of the Beck Depression Inventory-II for detecting major depressive episode in medical samples.

| Authors               | Sample                  | Cut-off | Sensitivity | Specificity | PPV   | NPV   | AUC   | % MDD | Criterion        |
|-----------------------|-------------------------|---------|-------------|-------------|-------|-------|-------|-------|------------------|
| Outpatients           |                         |         |             |             |       |       |       |       |                  |
| Arnarson et al. (41)  | Adult outpatients       | 20      | 82          | 75          | NR    | NR    | 87    | 42.1  | MINI             |
| Arnau et al. (42)     | Adult - primary care    | 18      | 94          | 92          | 54    | 99    | 96    | 23.2  | PHQ              |
| Beck & Gable 2001 (44)| Postpartum outpatients  | 20      | 56          | 100         | 100   | 93    | 95    | 12    | SCID-I           |
| Bunevicius et al. (45)| Coronary outpatients    | 14      | 89          | 74          | 29    | 98    | 90    | 11    | MINI             |
| Carney et al. (46)    | Insomnia outpatients    | 17      | 81          | 79          | NR    | NR    | 83.8  | NR    | SCID-I           |
| Chaudron et al. (48)  | Postpartum outpatients  | 20      | 45.3        | 91.1        | NR    | NR    | 90    | 37    | SCID-I           |
| Chilcot et al. (49)   | Renal hemodialysis      | 16      | 89          | 87          | 89    | 87    | 96    | 22.5  | MINI             |
| de Souza et al. (53)  | Huntington’s disease    | 11      | 100         | 66          | 48    | 100   | 85    | 50    | SCAN             |
| Dutton et al. (55)    | Adult - primary care    | 14      | 87.7        | 83.9         | 69.5  | 94.2  | 91    | 29.5  | PRIME-MD         |
| Frasure-Smith & Léspérance (58) | Coronary outpatients | 14      | 91.2        | 77.5         | NR    | NR    | 92    | 13.7  | SCID-I           |
| Jones et al. (63)     | Epilepsy outpatients    | 11      | 96          | 80          | 48    | 99    | 94    | 17.2  | MINI = SCID      |
|                       |                         | 15      | 84          | 87          | 55    | 97    | 92    | 6.7   | SCID-I           |
|                       |                         | 11      | 95.7        | 78.3         | 42    | 99    | 94    |       | MINI + SCID      |
| Hayden et al. (62)    | Obese bariatric         | 13      | 100         | 63.9         | 29.7  | 100   | 84.7  | 13.3  | SCID-I           |
| Pereira et al. (76)   | Pregnant outpatients    | 16      | 83.3        | 93.1         | 14.3  | 99.7  | 95    | 1.3   | DIGS             |
| Rampling et al. (78)  | Epilepsy outpatients    | 14      | 93.6        | 74           | 44    | 98    | 90    | 17.7  | MDI (ICD-10)    |
|                       |                         | 15      | 93.8        | 78.9         | 49.5  | 98    | 93    | 18    | MDI (DSM-IV)    |
| Su et al. (80)        | Pregnant outpatients    | 12      | 72.7-75.0   | 82.7-82.9    | NR    | NR    | 81.9-86.6 | 12.4 | MINI        |
| Tandon et al. (82)    | Perinatal women         | 12      | 84.4        | 81.0         | NR    | NR    | 91    | 33.7  | SCID-I           |
| Teng et al. (83)      | Postpartum outpatients  | 14      | 92          | 83           | 42    | 99    | 99    | 11.8  | MINI             |
| Turner et al. (84)    | Stroke outpatients      | 11      | 92          | 71           | NR    | NR    | 89    | 18    | SCID-I           |
| Turner-Stokes et al. (85)| Brain injury outpatients | 14     | 74          | 80           | 69    | 84    | NR    | 39.8  | DSM-IV          |
| Wan Mahmud et al. (87)| Postpartum outpatients  | 9       | 100         | 98           | 87.5  | 100   | 99.5  | 48    | CIS              |
| Warmenhoven et al. (88)| Cancer outpatients      | 16      | 90          | 69           | NR    | NR    | 82    | 22    | PRIME-MD         |
| Williams et al. (89)  | Parkinson outpatients   | 7       | 95          | 60           | 62    | 94    | 85    | 34.1  | SCID-I           |
| Hospital sample       |                         |         |             |             |       |       |       |       |                  |
| Homaifard et al. (94)| Traumatic brain injury  | 19      | 87          | 79           | NR    | NR    | NR    | 44.2  | SCID-I           |
| Huffman et al. (95)   | Myocardial infarction   | 16      | 88.2        | 92.1         | 62.5  | 98.1  | 96    | 13    | SCID-I           |
| Low & Hubbley (97)    | Coronary disease        | 10      | 100         | 75           | 21    | 100   | 92    | 11.8  | SCID-I           |
| Pietsch et al. (40)   | Adolescents             | 19      | 86          | 93           | 47    | 99    | 93    | 6.7   | Kinder-DIPS      |

**BDI-25:** Beck et al. (26) | Medical inpatients | 4 | 82 | 82 | NR | NR | 92 | 66 | PRIME-MD |
| Neitzer et al. (72)  | Renal hemodialysis     | 4      | 97.2        | 91.8         | 81.4  | 98.9  | 98   | 28.7  | BDI-II = 16    |
| Pietsch et al. (40)  | Adolescents            | 6      | 81          | 90           | 37    | 99    | 92   | 6.7   | Kinder-DIPS    |
| Poole et al. (103)   | Chronic pain outpatients | 4  | 81    | 92          | NR    | NR    | 94   | 59.4  | BDI-II = 19    |
| Scheinthal et al. (104)| Geriatric outpatients  | 5      | 75          | 93           | NR    | NR    | 94   | 47.8  | BDI-II = 22    |
| Steer et al. (107)   | Medical outpatients    | 4      | 97          | 99           | NR    | NR    | 99   | 24.2  | PRIME-MD        |
| Winters et al. (39)  | Adolescent outpatients | 4      | 91          | 91           | NR    | NR    | 98   | 11    | PRIME-MD        |

PPV: positive predictive value; NPV: negative predictive value; AUC: area under the curve; %MDD: proportion of major depression disorder; NR: not reported.

PHQ: PRIME-MD Patient Health Questionnaire; MINI: Mini International Neuropsychiatric Interview; PRIME-MD: Primary Care Evaluation of Mental Disorders; CIS: Clinical Interview Schedule; SCID-I: Structured Clinical Interview for DSM-IV Axis I Diagnosis; MDI: Major Depression Inventory; Kinder-DIPS: Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter; DIGS: Diagnostic Interview for Genetic Studies; SCAN: Schedules for Clinical Assessment in Neuropsychiatry.

---
Researchers have adopted both exploratory and confirmatory strategies with different purposes, e.g., to identify problems with items that have non-significant factor loadings or data cross-validation. The use of the state-of-the-art confirmatory approach is a trend in studies investigating the latent structure of BDI-II.

Using an exploratory strategy, Beck and colleagues reported a two-factor oblique structure for student and psychiatric samples (24), the cognitive-affective and somatic-vegetative dimensions. Although this bidimensional structure could be replicated among medical patients (30,42,43,50,54,56,75,77,86), several investigators reported different solutions (29,47,61,67,69,70,87). Somatic symptoms of depression have clustered as a dominant dimension, e.g., in primary care (42,86) and in coronary patients (54), or as an independent third dimension (29,61,67,69).

These alternative solutions could not be replicated by confirmatory strategy, but the somatic factor was observed as an ever-present factor among medical patients (Table 4). Summarizing the factor structure of the existing BDI investigations through meta-analysis (35), much of the data variability can be explained by the common dimension of “severity of depression” and by the other part, “somatic symptoms.” Due to the misattribution of somatic symptoms from medical conditions to depression, the assessment of depressive symptom severity with the BDI-II can be substantially biased in medically ill patients compared with non-medically ill patients. Among factor analytical investigations, the somatic dimension has emerged as being highly correlated with the cognitive dimension (>0.50, range 0.49-0.87).

The heterogeneous characteristics of depressive conditions could partially explain these proposed factor structures in medical patients. The alternative structural analysis of the BDI-II was strengthened by two model breakthroughs: the hierarchical model and the bifactor model. The hierarchical structure of higher-order depression to explain the variance of the lower-order cognitive and somatic dimensions was tested in several medical samples (42,54,56,61). Although scant, the bifactor model identified a scale solution with a general depression, in addition to the traditional bidimensional structure (50,101). The data variance of the BDI-II supported a higher order, or a parallel construct, of “general depression” and suggested caution when interpreting subscale scores.

**DISCUSSION**

The present systematic review is intended to aid practicing professionals and clinical researchers in several specialties in assessing depression in their patients and in interpreting the score through the BDI-II. Ideally, deciding which depression scale is optimal for use in medical settings should meet some desirable features from the patient’s and the clinician’s perspectives. Patients should find the measure user-friendly and the instructions easy to follow. The questions should be understandable and applicable to the patient’s problem. The scale should be brief to allow routine administration at intake and follow-up visits. From the clinician’s perspective, the instrument should provide clinically convenient information to increase the efficiency of medical evaluation. Clinicians should find the instrument user-friendly and easy to administer and score with minimal training. To be trustworthy, the information
Table 4 - Construct validity of the latent structure of the Beck Depression Inventory-II in medical samples.

| Study                        | Sample                  | Method         | Factor 1                   | Factor 2                   | Factor 3                   | Factor 4                   |
|------------------------------|-------------------------|----------------|----------------------------|----------------------------|----------------------------|----------------------------|
| **Normative study**          |                         |                |                            |                            |                            |                            |
| Beck et al. (24)             | College students        | EFA            | Cognitive-affective        | Somatic-vegetative         |                            |                            |
| Outpatient/Primary Care      |                         |                |                            |                            |                            |                            |
| Arnau et al. (42)            | Adult - primary care    | PCA            | Somatic-affective          | Cognitive                  | (Depression)               |                            |
| Brown et al. (43)            | Chronic fatigue        | EFA            | Cognitive                  | Somatic-affective          |                            |                            |
| Carvalho Bos et al. (47)     | Pregnancy outpatients   | PCA            | Cognitive-affective        | Anxiety                    | Fatigue                    |                            |
| Chilcot et al. (50)          | Renal disease outpatients | PCA                  | Cognitive-affective        | Fatigue                    |                            |                            |
| Corbière et al. (29)         | Chronic pain outpatients | EFA          | Cognitive                  | Affective                  | Somatic                    |                            |
| del Pino Pérez et al. (54)   | Coronary outpatients    | EFA            | Somatic-affective          | Cognitive                  |                            | (Depression)               |
| Grothe et al. (56)           | Adult - primary care    | CFA            | Cognitive                  | Somatic                    |                            |                            |
| Harris & D’Eon (61)          | Chronic pain outpatients | PCA            | Cognitive-affective        | Performance difficulty     | Somatic                    | (Depression)               |
| Kirsch-Darrow et al. (67)    | Parkinson outpatients   | CFA            | Cognitive-affective        | Loss of interest/pleasure  | Somatic                    |                            |
| Lipp et al. (69)             | HIV infection outpatients | C-PCA         | Cognitive                  | Affective                  | Somatic                    |                            |
| Lopez et al. (70)            | Chronic pain outpatients | EFA          | Negative rumination        | Somatic Complaint          | Mood                       |                            |
| Patterson et al. (75)        | Hepatitis C outpatients | CFA            | Cognitive-affective        | Somatic                    |                            |                            |
| Poole et al. (77)*           | Chronic pain outpatients | EFA          | Cognitive                  | Somatic-affective          | Behavior and activities    |                            |
| Viljoen et al. (86)          | Adult - primary care    | EFA            | Somatic-affective          | Cognitive                  |                            | (Depression)               |
| Wan Mahmud et al. (87)       | Postpartum outpatients  | PCA            | Affective                  | Somatic                    | Cognitive                  |                            |
| **Hospital sample**          |                         |                |                            |                            |                            |                            |
| Gorestein et al. (93)        | Adult - hospitalized    | EFA            | Cognitive-affective        | Somatic                    | Symptoms of depression     | Vegetative symptoms        |
| Rowland et al. (98)          | Traumatic brain injury  | PCA            | Negative self-evaluation   | Somatic                    |                            |                            |
| Siegert et al. (99)          | Neurological disease    | PCA            | Cognitive-affective        | Somatic                    |                            |                            |
| Thombs et al. (101)          | Acute myocardial infarction | CFA          | Cognitive                  | Somatic                    | General depression (G)     |                            |
| Tully et al. (102)           | Cardiac heart disease   | CFA            | Cognitive                  | Affective                  |                            |                            |

EFA: exploratory factor analysis; PCA: principal component analysis; C-PCA: confirmatory principal component analysis; CFA: confirmatory factor analysis.

(G) General factor of depression for the bifactor model.

(Depression) Higher order depression dimension for the hierarchical model.

* Only 18 items were used in the factorial model.

The Beck Depression Inventory-II (BDI-II) is a brief scale that is acceptable to patients and clinicians, covers all DSM-IV diagnostic criteria for major depressive disorder, and stands as a reliable indicator of symptom severity and suicidal thoughts. Its validity and case-finding capability as a screening instrument is well established. Conversely, its use as an indicator of sensitivity to change, medical patient’s remission status, psychosocial functioning, and quality of life deserve further investigation. The BDI-II is copyrighted and must be purchased from the publisher, which obstructs its wider use. Because direct comparisons demonstrating that the BDI-II is more reliable or valid than other depression scales are lacking, it is unwise to justify the cost of its systematic adoption.

Systematic reviews are susceptible to publication bias, that is the likelihood of over-representation of positive studies in contrast with non-significant results that frequently remain unpublished. In psychometric analyses due to its descriptive nature this kind of bias is minimized. Despite its reasonable psychometric characteristics, the BDI-II has some limitations. The spectrum bias refers to the differential performance of a test between different settings, thus affecting the generalizability of the results. For example, the somatic factor is a primary dimension among medical patients (42,54,86) instead of depressive cognition in non-clinical individuals. In addition, the work-up or verification bias occurs when respondents with positive (or negative) diagnostic procedure results are preferentially referred to receive verification by the gold-standard procedure, allowing considerable distortion in the accuracy of a given test. For example, medical patients with multiple somatic complaints might be routinely referred to psychiatric assessment and, thus, would be more likely labeled as depressed. To the extent that these types of bias may occur, the cut-off scores need to be checked psychometrically to convey the sample characteristics. Techniques assessing the item-level (e.g., item-total correlation and IRT analysis) and the scale-level (e.g., signal detection analysis and factor analysis) can improve the feasibility and strengthen the validity of using this scale to detect depressive symptoms in medical settings.

In the healthcare context, the perceived burden of scale completion by the clinician is the major obstacle to using standardized scales, such as the Hamilton Depression Rating Scale, which is unlikely to meet with success. As a self-report questionnaire to measure depression, the BDI-II holds the advantages of releasing the overburdened.
clinician from the paperwork of scale administration and of improving the efficiency of the clinical encounter by providing mental status assessment that correlates well with clinician-rated tools.

The stated purpose of the BDI-II is not to diagnose major depressive episode; thus, the investigators must grasp its appropriateness for detecting depressive symptoms and monitoring treatment efficacy and its comparability with observer-rated scales, such as the Hamilton Depression Rating Scale or the Montgomery-Åsberg Depression Rating Scale. Short scales that are less reliant on physical symptoms, such as the BDI-FS, should receive more investigation to demonstrate their usefulness in screening for depression in medically ill patients.

Finally, the BDI-II suffers from the intrinsic limitations of self-report questionnaires. Some individuals cannot complete the scale due to illiteracy, physical debility, or compromised cognitive functioning. The widespread use of the BDI-II among the elderly is not suggested. Reporting bias that minimizes or over-reports symptom severity is a possible hazard that reduces its validity in several patients.

As a tradeoff between the psychometric robustness and enumerated disadvantages of the BDI-II, this self-report scale can be viewed as a cost-effective option because it is sensitive to change. Br J Psychiatry. 1979;134:382-9.

Beck A, Rush A, Shaw B, Emery G. Cognitive therapy of depression. New York: Guildford Press; 1979.

Finally, the BDI-II suffers from the intrinsic limitations of self-report questionnaires. Some individuals cannot complete the scale due to illiteracy, physical debility, or compromised cognitive functioning. The widespread use of the BDI-II among the elderly is not suggested. Reporting bias that minimizes or over-reports symptom severity is a possible hazard that reduces its validity in several patients.

As a tradeoff between the psychometric robustness and enumerated disadvantages of the BDI-II, this self-report scale can be viewed as a cost-effective option because it is sensitive to change. Br J Psychiatry. 1979;134:382-9.

Beck A, Rush A, Shaw B, Emery G. Cognitive therapy of depression. New York: Guildford Press; 1979.

Finally, the BDI-II suffers from the intrinsic limitations of self-report questionnaires. Some individuals cannot complete the scale due to illiteracy, physical debility, or compromised cognitive functioning. The widespread use of the BDI-II among the elderly is not suggested. Reporting bias that minimizes or over-reports symptom severity is a possible hazard that reduces its validity in several patients.

As a tradeoff between the psychometric robustness and enumerated disadvantages of the BDI-II, this self-report scale can be viewed as a cost-effective option because it is sensitive to change. Br J Psychiatry. 1979;134:382-9.

Beck A, Rush A, Shaw B, Emery G. Cognitive therapy of depression. New York: Guildford Press; 1979.
53. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. J Clin Psychol. 2006;62(1):123-46, http://dx.doi.org/10.1002/jclp.20213.

54. Kozak K. A Psychology of Psychology: Oxford: American Psychological Association; 2010.

55. McDowell I. Measuring health: a guide to rating scales and questionnaires. 3rd ed. New York: Oxford University Press; 2006.

56. Dozois D. Beck Depression Inventory-II. In: Werner I, Craighead W, editors. The Corinri Encyclopaedia of Psychology. 4th ed. New York: John Wiley & Sons; 2010. p.210-1.

57. Winters JB, Steer RA, Jones-Hicks L, Beck AT. Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory-Second Edition for Primary Care. J Adolesc Health. 1999;24(6)::389-94, http://dx.doi.org/10.1016/S1054-1398(98)01359-9.

58. Pietsch K, Hoyer A, Fruehe B, Kruise J, Schulte-Korne J, Allgover AK. [Early detection of major depression in paediatric care: validity of the beck depression inventory-second edition (BDI-II) and the beck depression inventory-fast screen for medical patients (BDI-FS)]. Psychother Psychosom Med Psychol. 2012;62(11):418-24.

59. Amanson TO, Olsson DT, Smart J, Sigurethsson JF. The Beck Depression Inventory Second Edition (BDI-II): psychometric properties in Icelandic student and patient populations. Nord J Psychiatry. 62. Norway 2008.p. 360-5, http://dx.doi.org/10.1080/08039480801926681.

60. Amanu RC, Meagher MW, Norris MP, Bramson R. Psychometric evaluation of the Beck Depression Inventory-Second Edition in primary care medical patients. Health Psychol. 2001;20(2):112-9, http://dx.doi.org/10.1037/0278-1163.20.2.112.

61. Brown M, Kaplan C, Jason L. Factor analysis of the Beck Depression Inventory-Second Edition (BDI-II). J Psychopathol. Psychosom Med Psychol. 2012;62(11):418-24.

62. Carney CE, Ulmer C, Edinger JD, Krystal AD, Knauss F. Assessing depression symptoms in those with insomnia: an examination of the Beck Depression Inventory-Second Edition (BDI-II): do somatic symptoms compromise validity? J Psychosom Res. 2010;70(1):242-50, http://dx.doi.org/10.1016/j.jpsychores.2010.08.008.

63. Benevucis A, Stanute M, Brozaitiene J, Benevicius R. Diagnostic accuracy of self-rating scales for screening of depression in coronary artery disease patients. J Psychosom Res. 2012;71(1):122-5, http://dx.doi.org/10.1016/j.jpsychores.2011.10.006.

64. Carney CE, Ulmer C, Edinger JD, Krystal AD, Knauss F. Assessing depression symptoms in those with insomnia: an examination of the Beck Depression Inventory-Second Edition (BDI-II). J Psychopathol. Psychosom Med Psychol. 2012;62(11):418-24.

65. Kiropoulos LA, Meredith I, Tonkin A, Clarke D, Antonis P, Plunkett J. Psychometric properties of the cardiac depression scale in patients with coronary heart disease. BMC Psychiatry. 2012;12:216, http://dx.doi.org/10.1186/1471-244X-12-216.

66. King PR, Donnelly KT, Donnelly JP, Dunnam M, Warner G, Kittleson PR, et al. The MG-QOL15 Japanese version: validation and associations with clinical factors. Muscle Nerve. 2012;46(2):166-73, http://dx.doi.org/10.1002/mus.23398.

67. Kirsch-Darrow L, Marsiske M, Okun MS, Bauer R, Bowers D. Apathy and depression: separate factors in Parkinson’s disease. J Int Neuropsychol Soc. 2011;17(6):1058-66, http://dx.doi.org/10.1017/S1355617711000950.

68. Ko PW, Hwang J, Lim HW, Park SP. Reliability and validity of the Korean version of the Neurological Disorders Depression Inventory for Epilepsy (K-NDDI-E). Epilepsy Behav. 2012;25(4):539-42, http://dx.doi.org/10.1016/j.yebeh.2012.09.010.

69. Lips GE, Love GA, De La Haye W, Longman-Mills S, Clarke TR, Barton EN, et al. Validation of the Beck Depression Inventory II in HIV-positive Patients. West Indian Med J. 2010;59(4):374-9, http://dx.doi.org/10.1111/j.1533-2500.2010.00274.x.

70. Lopez MN, Pierce RS, Gardner RD, Hanson RW. Standardized Beck Depression Inventory-II Scores for Male Veterans Coping With Chronic Pain. Psychol Serv. 2013;10(2):257-63, http://dx.doi.org/10.1037/a0027920.

71. Masuda M, Utsugisawa K, Suzuki S, Nagane Y, Kabasawa C, Suzuki Y, et al. The MG-QOL15 Japanese version: validation and associations with clinical factors. Muscle Nerve. 2012;46(2):166-73, http://dx.doi.org/10.1002/mus.23398.

72. Neitzner A, Sun S, Doss S, Moran J, Schiller B. Beck Depression Inventory-Fast Screen (BDI-FS): an efficient tool for depression screening in patients with end-stage renal disease. Hemodial Int. 2012;16(2):207-13, http://dx.doi.org/10.1111/j.1542-4758.2012.00663.x.

73. Oeoms E, Meganck R, Vanhuele V, Binck W, Wateliet J, Dhoooge I. Tinnitus severity and the relation to depressive symptoms: a critical study. Otolaryngol Head Neck Surg. 2011;145(2):276-81, http://dx.doi.org/10.1177/019459981140381.

74. Osada K, Oka H, Isomura T, Nakamura I, Tominaga K, Takahashi S, et al. Development of the Japanese version of the Fibromyalgia Impact Questionnaire (FIQ): psychometric assessments of reliability and validity. Int J Rheum Dis. 2012;15(1):74-81, http://dx.doi.org/10.1111/j.1609-6314.2011.00892.x.

75. Patterson AL, Morasco BJ, Fuller BE, Indest DW, Loftis JM, Hauser P. Screening for depression in epilepsy clinics. A comparison of routine intake interview versus conventional and visual-analog methods. Epilepsia. 2012;53(10):1713-9, http://dx.doi.org/10.1111/j.1528-1167.2012.03571.x.

76. Poole H, White S, Blake C, Murphy P, Bramwell R. Depression in chronic pain patients: prevalence and measurement. Pain Pract. 2010;10(6):479-84, http://dx.doi.org/10.1111/j.1533-2500.2009.00274.x.

77. Rampilng J, Mitchell AJ, Von Oertzen T, Dockcr J, Jackson C, Cock H, et al. Screening for depression in epilepsy clinics. A comparison of conventional and visual-analog methods. Epilepsia. 2012;53(10):1713-21, http://dx.doi.org/10.1111/j.1528-1167.2012.03371.x.

78. Rampilng J, Mitchell AJ, Von Oertzen T, Dockcr J, Jackson C, Cock H, et al. Screening for depression in epilepsy clinics. A comparison of conventional and visual-analog methods. Epilepsia. 2012;53(10):1713-21, http://dx.doi.org/10.1111/j.1528-1167.2012.03371.x.
Beck Depression Inventory-II in medical patients
Wang Y-P and Gorenstein C

97. Rowland SM, Lam CS, Leahy B. Use of the Beck Depression Inventory-II (BDI-II) with persons with traumatic brain injury: Analysis of factorial structure. Brain Inj. 2005;19(2):77-83, http://dx.doi.org/10.1080/ 02699050410001719988.
98. Stenlert RJ, Walkely FH, Turner-Stokes L. An examination of the factor structure of the Beck Depression Inventory-II in a neurorehabilitiation inpatient sample. J Int Neuropsychol Soc. 2009;15(1):142-7, http://dx.doi.org/10.1017/S1355617708090408.
99. Thomas SA, Lincoln NB. Depression and cognitions after stroke: validation of the Stroke Cognitions Questionnaire Revised (SCQR). Disabil Rehabil. 2008;30(23):1779-85, http://dx.doi.org/ 10.1080/09638280701661430.
100. Thombs BD, Ziegelstein RC, Beck CA, Pilotte L. A general factor model for the Beck Depression Inventory-II: validation in a sample of patients hospitalized with acute myocardial infarction. J Psychosom Res. 2008;65(2):115-21, http://dx.doi.org/10.jpsychores.2008.02.027.
101. Tully PJ, Wineslief HR, Baker RA, Turnbull DA, de Jonge P. Confirmatory factor analysis of the Beck Depression Inventory-II and the association with cardiac morbidity and mortality after coronary revascularization. J Health Psychol. 2011;16(4):584-85, http://dx.doi.org/ 10.1177/1359105310386064.
102. Poole H, Bramwell R, Murphy P. The utility of the Beck Depression Inventory Fast Screen (BDI-FS) in a pain clinic population. Eur J Pain. 2009;13(8):865-9, http://dx.doi.org/10.1016/j.ejpain.2009.09.017.
103. Schenfield SM, Steer RA. Evaluating geriatric medical outpatients with the Beck Depression Inventory-Fast Screen for medical patients. Aging Ment Health. 2001;5(2):143-8.
104. Servaes P, van der Werf S, Prins J, Verhagen S, Bieijenberg F. Fatigue in chronic-free cancer patients compared with fatigue in patients with Chronic Fatigue Syndrome. Support Care Cancer. 2001;9(1):11-7, http://dx.doi.org/10.1007/s0052000165.
105. Servaes P, Prins J, Verhagen S, Bieijenberg F. Fatigue after breast cancer and in chronic fatigue syndrome - Similarities and differences. J Psychosom Res. 2002;52(6):453-9, http://dx.doi.org/10.1016/s0022- 3992(02)00320-6.
106. Steer RA, Cavaliere TA, Leonard DM, Beck AT. Use of the Beck Depression Inventory for Primary Care to screen for major depression disorders. Gen Hosp Psychol. 1999;21(2):106-11, http://dx.doi.org/10.1016/s0163-4803(99)00070-x.
107. Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
108. Hanel G, Henningsen P, Herzog W, Sauer N, Schaeffer N, Szeleszyni et al., Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study. J Psychosom Res. 2009;67(3):189-97.
109. Voigt K, Nægå A, Meyer B, Langs G, Braukhaus C, Löwe B. Towards positive diagnostic criteria: A systematic review of somatoform disorder diagnoses and suggestions for future classification. J Psychosom Res. 2010;68(5):405-14, http://dx.doi.org/10.1016/j.jpsychosomres.2010.01.015.
110. Ramasubbu R, Beaulieu S, Taylor VH, Schaffer A, McIntyre RS. The CANMAT task force recommendations for the management of patients with somatoform disorders and comorbid medical conditions: diagnostic assessment, and treatment principles. Ann Clin Psychiatry. 2012;24(1): 82-90.
111. Zgigmond AS, Smith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67(6):261-70, http://dx.doi.org/10.1111/j.1600- 0447.1983.tb00976.x.
112. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale - An updated literature review. J Psychosom Res. 2002;53(2):69-77, http://dx.doi.org/10.1016/s0022-3991(01)00296-3.
113. Cosco TD, Doyle F, Ward M, McGee H. Latent structure of the Hospital Anxiety And Depression Scale: A 10-year systematic review. J Psychosom Res. 2012;73(2):180-4, http://dx.doi.org/10.1016/j.jpsychores.2011.06.008.
114. Chassamzadeh H, Motalab A, Karamghadiri N, Ebrahimkhani N. Psychometric properties of a Persian-language version of the Beck Depression Inventory second edition: BDI-II-Persian. Depression Anxiety. 2012;25(1):185-92, http://dx.doi.org/10.1002/da.21702.
115. Sprinkle SD, Lurie D, Insko SL, Atkinson G, Jones GL, Logan AR et al. Criterion Validity, Severity Cut Scores, and Test-Retest Reliability of the Beck Depression Inventory-II in a University Counseling Center Sample. J Couns Psychol. 2012;69(2):381-3.
116. Hambleton R, Swaminathan H, JH R. Fundamentals of item response theory. New York: Macmillan; 1993.
117. Waller NG, Combes PA, Bullon SD, Beckford J. Measurement of depressive symptoms in women with breast cancer and women with clinical depression: A differential item functioning analysis. J Clin Psychol Med Settings. 2005;12(2):221-4, http://dx.doi.org/10.1080/ 10863280500343727.
118. Leonardson GR, Kemper E, Ness FK, Koplin BA, Daniels MC, Leonardson GA. Validity and reliability of the audit and CAGE-AID in Northern Plains American Indians. Psychol Rep. 2005;97(1):161-6.
119. Anastasi A, Urbina S. Psychological Testing. 7th ed. London: Prentice-Hall International; 1997.
121. Shean G, Baldwin G. Sensitivity and specificity of depression questionnaires in a college-age sample. J Genet Psychol. 2008;169(3):281-8.
122. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton depression rating scale: Has the gold standard become a lead weight? Am J Psychiatry. 2004;161(12):2163-77.
123. Kriston L, von Wolff A. Not as golden as standards should be: Interpretation of the Hamilton Rating Scale for Depression. J Affect Disord. 2011;128(1):175-7, http://dx.doi.org/10.1016/j.jad.2010.07.011.
124. Cicchetti DV. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. Psychol Assess. 1994;6(4):284-90, http://dx.doi.org/10.1037/1040-3590.6.4.284.
125. Moran P, Lambert M. A review of current assessment tools for monitoring changes in depression. In: Lambert M, Christensen E, DeJulio S, editors. The Assessment of Psychotherapy Outcome. New York: Wiley; 1983.
126. Osman A, Barrios FX, Gutierrez PM, Williams JE, Bailey J. Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. J Clin Psychol. 2008;64(1):83-102, http://dx.doi.org/10.1002/jclp.20433.
127. Byrne BM. Factor analytic models: viewing the structure of an assessment instrument from three perspectives. J Pers Assess. 2005;85(1):17-32, http://dx.doi.org/10.1207/s15327752apa8501_02.
128. Byrne B. Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications and programming. Mahwah, New Jersey: Lawrence Erlbaum Associates; 1998.