

Discriminability of the Beck Depression Inventory and its Abbreviations in an Adolescent Psychiatric Sample

Fatemeh Seifi^{1,*}, Sebastian Therman², Tommi Tolmunen^{1,3}

¹ Clinical Medicine Unit, Department of Medicine/Adolescent Psychiatry, University of Eastern Finland, Kuopio, Finland

² Mental Health, Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland

³ Department of Adolescent Psychiatry, Kuopio University Hospital, Kuopio, Finland

*Corresponding author: fseifi@uef.fi

Abstract

Background: The Beck Depression Inventory (BDI) is a widely acknowledged self-report screening tool for evaluating the presence and intensity of depressive symptoms. The BDI-IA, although an older version, is highly correlated with the updated BDI-II, remains clinically valuable, and is widely used due to its free availability.

Aim: This study aimed to examine the psychometric properties of the BDI-IA and compare its diagnostic accuracy with the abbreviated BDI-SF, BDI-PC, and BDI-6 versions against gold-standard research diagnoses in a representative Finnish adolescent clinical population.

Methods: The participants were referred outpatient adolescents aged 13–20 years ($N = 752$, 73% female). We investigated structural validity with item factor analysis and evaluated the criterion validity of mean scores and factor scores with various diagnostic measures. Sample-optimal cut-offs (criterion unweighted Cohen's kappa) were estimated with a bootstrap procedure.

Results: The sample-optimal cut-off for the full BDI was 19, slightly higher than that suggested by the previous literature. The abbreviations of the BDI-IA were demonstrated to be as good as the full scale in detecting depressive symptoms in all three diagnostic categorizations.

Conclusion: The use of brief and user-friendly questionnaires such as the BDI-PC or BDI-6 is recommended to ensure optimal depression screening and minimize the administrative burden, especially in primary care settings where clinical decision-making and referrals often need to occur within a limited time frame.

Keywords: Adolescent, depression, screening, factor analysis, diagnostic accuracy

Introduction

In the past two decades, depression in adolescents has been a growing mental health concern worldwide due to its profound long-term impacts on psychosocial and academic functioning (1). In clinical settings, major depressive disorder (MDD) in adolescents is an especially important issue due to its high incidence and the high probability of exacerbation of untreated symptoms persisting into adulthood, and the increased risk of suicidal behaviors (2). In addition, during the COVID-19 pandemic, adolescent depression approximately doubled globally compared to the pre-pandemic level (from 12.9% to 25.2%) (3). According to WHO, Finland, for example, witnessed a substantial increase in depressive symptoms among adolescents, with

17–19% of 15-year-old females reporting feeling down in their everyday lives in 2022 (4).

It is well-established that valid screening instruments could be beneficial, particularly in clinical settings where symptomatic adolescents are seen in primary care services within a constrained time frame (5). Furthermore, the early detection of depressive symptoms can facilitate the appropriate referral and admission of adolescent outpatients to secondary care, hence accelerating the treatment process and preventing later harmful effects (5).

One of the most widely acknowledged self-report screening tools for assessing the presence and intensity of depressive symptoms is the Beck Depression Inventory (BDI) (6). It has been noted that one specific strength of BDI's factor structure is its expanded format, whereby complete sentences are

used instead of rating scales such as the Likert format, thus avoiding the potential confusion, errors, and carelessness of respondents caused by negative reverse-worded items (7), as well as subjective interpretations of response options such as sometimes or often. The original 21-item version of the BDI has undergone several modifications since it was first published in 1961. In the first revised version, known as the BDI-IA, the wording of several items was altered (8). Subsequently, in the upgraded or latest version (i.e., BDI-II (9)), some items have been updated to correspond with the criteria for MDD in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (10). The BDI-IA, although an older version, is still of clinical value and widely utilized due to its public availability and free use (11). Additionally, the Finnish version of the BDI-IA is widely accessible through Finnish medical and nursing databases, which further facilitates its use among clinicians and researchers in Finland (12). Moreover, previous studies have reported reasonable psychometric properties and high levels of internal consistency for the earlier BDI versions across different settings (13,14), and even with the revisions, the BDI-IA and BDI-II have demonstrated a strong correlation ($r = 0.93$) (15).

Despite the wide variety of literature evaluating the factor structure of the BDI in general and clinical populations across different countries (16,17), most validity reports, including those in Finland, have addressed the adult population (18,19). The results of a systematic review suggest that the factor structure of the BDI for adolescents, particularly in clinical contexts, could be different from that for nonclinical adolescents and adults (20). It is unclear, however, whether some degree of multidimensionality impacts the use of the measure for screening. Furthermore, essential unidimensionality presents an opportunity for abbreviation by removing the least discriminative items. While the 21-item BDI demonstrates a broad content coverage of depressive symptoms (21), the use of brief screening questionnaires, such as the abbreviated versions with 13 items (BDI-SF) (22), 7 items (BDI-PC) (23), or 6 items (BDI-6) (24), may thus make it more user-friendly in a clinical context without significant loss of accuracy (25).

Various BDI thresholds for suspecting the presence of depressive disorders and the need for referral have been suggested in studies with adolescent samples. Nevertheless, there is a lack of consensus regarding whether the provided optimal cut-off values can be generalized to clinical use among adolescents. In this respect, a recent meta-analysis established that BDI scores of both 11 and 16 could yield favorable diagnostic accuracy for MDD in adolescents, although the latter provided

slightly better detection (26). There are also several studies that report BDI threshold values for the clinical settings other than the above (27–29).

Relatively little research has been undertaken to test the validity of the BDI-IA or the abbreviated versions of the BDI in adolescent psychiatric outpatients or to set thresholds for their use. One study demonstrated that the brief self-assessment BDI-6 had satisfactory psychometric validity in an adolescent psychiatric context (25). However, this abbreviated format did not incorporate the crucial suicidality item. Hence, the present study aims to fill the above gaps by examining and comparing the diagnostic accuracy, validity, and psychometric properties of the BDI-IA, BDI-SF, BDI-PC, and BDI-6 questionnaires against the gold-standard Structured Clinical Interview for DSM Disorders (SCID) in a representative Finnish adolescent clinical population. Furthermore, we aimed to apply receiver operating characteristics (ROC) curve analysis to propose optimal cut-off scores and a sample-optimized BDI abbreviation and to establish generalizable data enabling the more effective and accurate detection of depressive symptoms among adolescents in clinical settings. Note that the technical term ‘diagnostic accuracy’ widely used in the field does not imply that the questionnaires are used to determine diagnoses, but for screening.

Methods

Participants

The data for the current investigation were derived from the ongoing REAL-SMART project (“Recognition and early intervention for alcohol and substance abuse in adolescence and systemic metabolic alterations related to different psychiatric disease categories in adolescent outpatients”). The participants were referred patients aged 13–20 years who attended the adolescent psychiatric outpatient clinic of Kuopio University Hospital (KUH) in Finland between June 2017 and March 2022, with breaks in clinic operations due to the COVID-19 pandemic. The reasons for non-participation were not recorded but included declining to participate, dropping out or being transferred to inpatient treatment before being approached, and not being presented with the study by clinical personnel for various reasons. Of the adolescents having at least one appointment in the outpatient clinic ($n = 2853$), a total of 754 (26.4%) participants were enrolled in the study, of whom the majority (73 %) were female, which is a typical gender distribution for these clinics. Previous or current diagnoses did not affect recruitment. Two patients were excluded for not filling in the BDI-IA at baseline.

The participants were first interviewed and then autonomously completed a multi-measure

questionnaire containing the BDI-IA on a tablet computer later in the same session. All items were presented one at a time, with automatic advancement after giving a response. Although it was possible to return to change earlier responses by swiping right on the screen, this functionality was not advertised, and only a few respondents used it for only a few responses.

Prior to undertaking the investigation, the Research Ethics Committee of Kuopio University Hospital confirmed the study procedures and written informed consent was obtained from all the participants. In addition, the study complied with the ethical principles set by the 7th revision of the Declaration of Helsinki (30).

To enhance the transparency of the present research, we adhered to the STARD 2015 (Standards for the Reporting of Diagnostic Accuracy Studies) checklist (Supplementary material Table S1), which consists of 30 items (31).

Measures

Research diagnoses

A trained psychiatric nurse conducted a Structured Clinical Interview for DSM-IV, clinician version (SCID-CV), with each participant (32), assigning all comorbid diagnoses. The involvement of a trained psychiatric nurse was in accordance with the SCID user's guide, which permits structured diagnostic interviews and the assignment of diagnoses by trained professionals who are not necessarily psychiatrists or psychologists (32).

The interviewer was blind to BDI responses when completing the SCID and assigning research diagnoses. The following diagnoses were exclusion criteria in discriminability and optimal cut-off analyses, as the presence of depressive symptoms independent of depression diagnoses were undetermined for them: all psychotic disorders except Major Depressive Disorder (MDD) with Psychotic Features, Bipolar Disorder, Cyclothymic Disorder, and mood disorders other than MDD or Dysthymia. The SCID-based diagnoses are presented in Table 1. To ensure the robustness of our results, we dichotomized the presence or absence of depression diagnoses in three different ways, with one assigned as the main classification. In the primary categorization, MDD vs. no depression, patients with major depressive disorder (MDD), either current or in partial remission, were coded as depressed, patients with dysthymia were excluded, and all others were coded as not depressed. Note that MDD in full remission was coded as not depressed. The secondary categorization, depression vs. none, was otherwise the same, except that participants with dysthymia were also coded as depressed. In the other secondary categorization, MDD vs. no MDD, those

with dysthymia (without MDD) were coded as no MDD. Note that the depressed cases were the same for the primary categorization MDD vs. No Depression and the secondary MDD vs. no MDD, and the non-depressed cases were the same for the primary categorization and the secondary Depression vs. none.

Beck Depression Inventory I-A (BDI-IA)

Participants completed the self-report 21-item BDI-IA (8), which was previously translated into Finnish (12). Each item of the BDI-IA and its abbreviations is scored from 0 to 3, with a sum score range from 0 to 63 for the BDI-IA (see the items in Table 3). We used the sum score screening thresholds (primarily 16 or greater and secondarily 11 or greater) recommended in a recent meta-analysis (26). The BDI-IA and its abbreviations were scored automatically and thus blind to the results of the SCID.

Beck Depression Inventory – Short Form (BDI-SF)

The BDI-SF (22) consists of 13 items extracted from the original BDI to evaluate the presence of the depressive symptoms. This set of optimal items has previously revealed high correlations with the total score for the original version (22,33).

Beck Depression Inventory for Primary Care (BDI-PC)

The BDI-PC, also referred to as the BDI-FS ("Fast Screen"), is composed of 7 non-somatic items of the BDI for screening MDD in primary care patients (23). Previous studies have reported high internal consistency for the BDI-PC in adult medical inpatients and outpatients (23,34) and adolescent outpatients in primary care (35).

Beck Depression Inventory-6 (BDI-6)

Two brief scales, named BDI-6, comprise six items extracted from the BDI-21 and BDI-IA. The selected items differ in the validation studies we found (24,25,36), and the suicidality item is not included in either. Both BDI-6 versions demonstrated acceptable criterion validity in the studies mentioned above. In the current study, we investigated the BDI-6 version presented by Blom et al. (2012) (25), as it was based on the BDI-IA rather than the original BDI-21.

Statistical analyses

All statistical analyses were performed in the R software environment (version 4.4.0) (37). When calculating BDI sum scores, missing BDI responses were replaced with the participant's mean for the other responses of the respective form version. Overall, missingness was minimal (0.5% of all responses).

For the nonparametric comparison of distributions, we used the stochastic superiority index, $p^{\wedge} = (X > Y) + 0.5 \times (X = Y)$, also known as the common language effect size, which indicates the probability that a random member of subgroup X has a higher score than a random member of subgroup Y. To test the statistical significance of p , we used the appropriate Brunner–Munzel test (38,39), implemented in the *brunnermunzel* R package (version 2.0) (40). All statistical tests were considered significant at $p < 0.05$ unless stated otherwise.

Sample size calculations

The sample was not collected specifically to test the BDI's diagnostic accuracy, and this study was conceived after the data had been gathered. Thus, data collection was not informed by the present study's needs. However, previous studies on the single-test accuracy of the BDI (26) had a median sample size of 316, indicating that our study was adequately powered (see Table S2 for detailed information).

We estimated the required sample size for comparing two binary diagnostic tests (the pre-specified cut-off applied to two BDI versions) in a paired design with the calculations suggested by Akoglu (2022) (41), with a Type I error rate (α) of 5%, 95% power, and conservatively applying Yates' continuity correction. For detecting a clinically meaningful difference of .05 in either sensitivity or specificity between versions, the required sample size is 249 when the test agreement is at maximum and 1908 at minimum. Since the compared BDI versions were various summations of nearly the same responses, test agreement can be expected to be close to the maximum, and the present sample size was, therefore, more than sufficient.

As we secondarily report differences in BDI distributions between the depressed and non-depressed groups, we also estimated the required sample size for these comparisons. We are not aware of a specific power calculation formula for the Brunner–Munzel test, but it is at least as powerful as the Wilcoxon–Mann–Whitney U-test in most situations (39). We, therefore, estimated the two-sided power of the U-test with the formulas of Noether (1987) (42), as implemented in the *rankFD* package (version 0.1.1) (43) for R. At an α of 5% and 95% power, detecting even a modest effect of $p^{\wedge} = 0.6$ required only 217 participants each in the two compared groups. The sample was thus clearly sufficient, even though the estimation was slightly optimistic in not accounting for ties.

Factor analysis

The latent structure of the BDI-IA was investigated with a confirmatory factor analysis of the a priori

single-dimensional model, treating the items as ordinal (DWLS estimator), with the *cfa* function in the *lavaan* package (version 0.6-17) (44), using all pairwise-available data and otherwise standard settings. Factor scores were computed with the *lavPredict* function using the default empirical Bayes modal approach. Factor scores for the abbreviated forms were calculated using the parameters from the full model, thus keeping parameters identical across the form versions and conceptually treating missing items in an abbreviated version as missing responses of the full BDI.

Discriminability

Basic accuracy ratios, Cohen's kappa (κ), the diagnostic odds ratio (DOR), and the number needed to diagnose (NND; reciprocal of the Pierce skill score, a.k.a. Youden's J index) for each BDI version in detecting gold-standard depression diagnoses were computed with the *epi.tests* function of the *epiR* package (version 2.0.74) (45). The NND can here be interpreted as the number of individuals that need to be screened to correctly detect one person with a diagnosis (46). The primary full-length BDI cut-off of 0.76 for these analyses was the mean score equivalent to a sum of 16 or greater, as described above, and the secondary mean score cut-off was 0.52. For the abbreviated forms, apparent prevalences were higher at these cut-offs due to higher scores for the included items. For comparability, we therefore matched cut-offs with the closest equivalent in sensitivity; when two cut-off candidates were equal in their sensitivity difference, the cut-off with the smallest difference in specificity was selected.

To compare the diagnostic performance of the BDI-IA with the BDI-SF, BDI-PC, and BDI-6 in a paired design, the discriminability of the mean scores of the shortened forms was compared with those of the full 21-item form using the statistical procedure of Roldán-Nofuentes (2020) (47) implemented in the R software package *testCompareR* (version 1.0.3) (48). First, a global test jointly compared sensitivity and specificity for detecting the gold-standard diagnosis, and if this test was statistically significant, sensitivity and specificity differences were tested separately. Multiple tests within the same form pair were corrected with Holm's method, the package default. Due to cut-off matching by sensitivity, this procedure mainly compared specificity but considered the sensitivity discrepancies arising from imperfect matching due to coarse sum score distributions.

As sum scores have a simpler measurement model, which is potentially less accurate than factor scores, we compared the diagnostic categorization by factor scores with categorization by the full BDI-IA sum

score, in the same manner as the comparisons with the mean scores of the abbreviations. Again, the factor score cut-off was selected to match the sensitivity of the sum score cut-off, and accuracy was compared with testCompareR as above. The strength of association between the binary categorizations by mean scores and factor scores was expressed as a correlation (equivalent to Stuart's τ_c).

Optimal cut-offs

As an exploratory addition to the main results, we determined sample-optimal cut-offs for all the BDI

versions using the R package cutpointr (version 1.1.2) (49) with the bootstrapped version of the cutpointr function using default settings, the chosen cut-off being validated in 1000 out-of-bag samples. The main criterion for comparison was Cohen's unweighted κ , estimated jointly and for genders separately, with secondary criteria being misclassification cost with a) false positives and false negatives being weighted equally or b) with triple weight on the latter. Discriminability across the whole score range of each form was assessed with

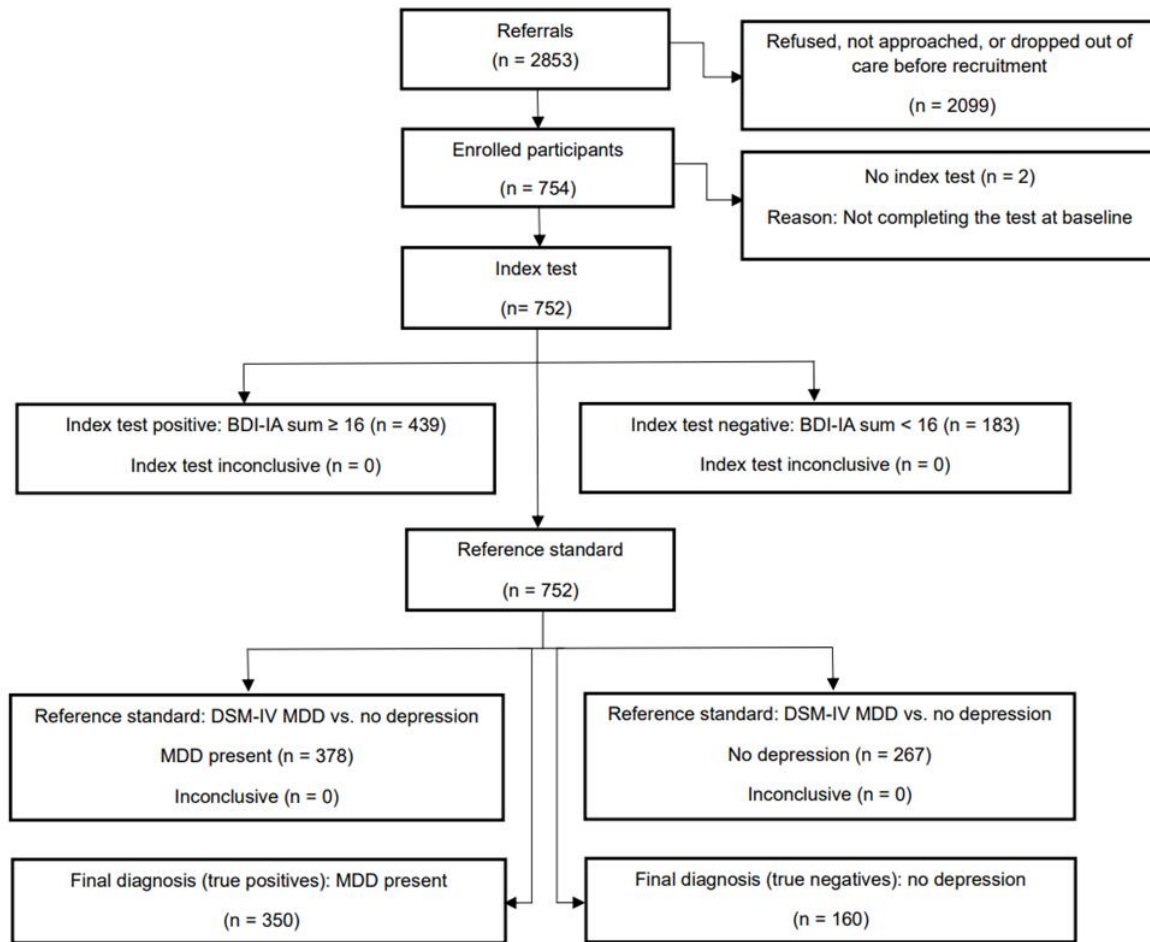


FIGURE 1. Flow diagram of participants based on the primary diagnostic categorization (MDD vs. no depression). Note: MDD vs. no depression diagnosis refers to the main comparison categorization used in the analysis. For more details, see Methods section: Research Diagnoses.

TABLE 1. Characteristics and presence of SCID-based diagnoses by gender.

Variable	Males	Females	Total
Participants	202	550	752
Age, mean	16.74	16.47	16.55
Age, SD	1.67	1.62	1.64
Major depressive disorder	84 (42%)	294 (53%)	378 (50%)
Dysthymia	24 (12%)	87 (16%)	111(15%)
Other depressive disorder	7 (3.5%)	29 (5%)	36 (5%)
Exclusion diagnosis	6 (3%)	61 (11%)	67 (9%)

SCID = Structured Clinical Interview for DSM-IV; SD = Standard deviation. Note: an individual may have diagnoses in more than one category. Exclusion diagnoses: psychotic disorders other than major depressive disorder (MDD) with psychotic features, bipolar disorder, and cyclothymic disorder

TABLE 2. BDI sum score distributions by version and diagnostic category.

Condition*	Version	Diagnosis +			Diagnosis -			Brunner–Munzel test		Stochastic Superiority	
		n	M	SD	n	M	SD	Statistic	p	Estimate	CI
MDD vs. no depr.	BDI-1A	351	30.22	9.64	239	12.13	9.49	31.7	< 0.001	0.902	[0.88, 0.93]
MDD vs. no depr.	BDI-SF	359	20.24	6.60	243	7.75	6.56	32.3	< 0.001	0.902	[0.88, 0.93]
MDD vs. no depr.	BDI-PC	361	11.58	3.91	245	4.54	3.91	29.5	< 0.001	0.892	[0.87, 0.92]
MDD vs. no depr.	BDI-6	368	10.02	3.04	246	4.00	3.29	32.7	< 0.001	0.899	[0.88, 0.92]
Depr. vs. none	BDI-1A	389	29.13	10.04	239	12.13	9.49	28.3	< 0.001	0.884	[0.86, 0.91]
Depr. vs. none	BDI-SF	397	19.53	6.87	243	7.75	6.56	28.7	< 0.001	0.885	[0.86, 0.91]
Depr. vs. none	BDI-PC	399	11.20	4.03	245	4.54	3.91	26.8	< 0.001	0.876	[0.85, 0.90]
Depr. vs. none	BDI-6	406	9.74	3.12	246	4.00	3.29	29.8	< 0.001	0.886	[0.86, 0.91]
MDD vs. no MDD	BDI-1A	351	30.22	9.64	275	13.04	9.55	30.7	< 0.001	0.892	[0.87, 0.92]
MDD vs. no MDD	BDI-SF	359	20.24	6.60	279	8.42	6.63	30.9	< 0.001	0.891	[0.87, 0.92]
MDD vs. no MDD	BDI-PC	361	11.58	3.91	281	4.95	3.96	28.1	< 0.001	0.880	[0.85, 0.91]
MDD vs. no MDD	BDI-6	368	10.02	3.04	282	4.40	3.35	30.2	< 0.001	0.883	[0.86, 0.91]

MDD = Major Depressive Disorder; BDI-1A = Beck Depression Inventory (revised version); BDI-SF = Beck Depression Inventory – Short Form; BDI-PC = Beck Depression Inventory for Primary Care; BDI-6 = Beck Depression Inventory-6; MDD = Major Depressive Disorder; Depr = Depression; M = Mean; SD = Standard deviation; P = P-value; CI = Confidence interval. Note: Missingness of BDI responses was minimal, with only 85 responses of the 752 respondents missing (0.5%). Note: *Condition refers to the comparison categorizations used in the analysis. For more details, see Methods section 2.2.1: Research Diagnoses.

receiver operating characteristic (ROC) curves and accompanying area under the curve (AUC) values, as recommended in the STARD 2015 guidelines (50). For details of the interpretation of AUC values, see Mallet et al. (2012) (51).

Results

A participant flow diagram is provided in Figure 1. The demographic and diagnostic distributions of the participants are presented in Table 1. BDI score distributions and comparisons thereof between the diagnostic groups are reported in Table 2. As expected, there was little overlap in BDI scores between those with a diagnosis of depression and those without, ρ being around 0.9 across diagnostic groups and BDI versions, which corresponds to a standardized mean difference of 1.8.

Factor analysis

The fit of the one-dimensional factor model of the BDI-1A was sufficient, as the scaled comparative fit index (CFI) was 0.949, the scaled root mean square error of approximation (RMSEA) was 0.089, and the standardized root mean square residual (SRMR) was 0.060. Items 19 Weight loss and 20 Somatic preoccupation had the weakest factor loadings (0.24 and 0.42, respectively), while the core items 1, 3, 4, 5, 7, and 8 pertaining to depressive mood and negative self-view had loadings of 0.80 or above. Thresholds (standard scores corresponding to the cumulative response probability within an item) had a wide range: for example, the highest score of 3 on item 10 Crying was as frequent as the score of 1 on item 19 Weight loss. See the standardized parameters of the factor model in Table 3 and BDI factor score distributions by diagnostic category in Table S3.

Discriminability

The main screening discriminability results are presented in Table 4 and Table 5. The primary cut-off (16 or higher on the full BDI) turned out to be relatively low in this highly symptomatic sample, with a sensitivity of 0.94 but a specificity of only 0.64. In the global test, the mean scores of all the abbreviated versions were as good at discriminating between MDD and no depression as the full BDI-1A. Although the differences were not statistically significant, the BDI-6 had a higher negative predictive value (NPV) than the BDI-1A at the same positive predictive value (PPV), which was also reflected in having the lowest NND and a higher DOR. The 7-item BDI-PC was the only abbreviation to have a lower α and worse NND than the BDI-1A, although, again, these differences were not statistically significant, and the DOR was the highest of all the versions. These differences were largely attributable to a higher threshold due to a lack of perfectly matched cut-off values. The 13-item BDI-SF performed nearly identically to the BDI-1A. The results of the robustness analysis using the two alternate diagnostic categorizations were largely the same (Tables S4 and S5).

At the secondary cut-off of 11 or higher, the diagnostic accuracy of the form versions did not differ, probably due to the sensitivities being extreme (0.98 for the BDI-1A) at such a low cut-off (Tables S6 and S7). ROC curves for the BDI-1A version are displayed in Figures 2a and 2b, and all the form versions are provided in Supporting information Figure S1.

Factor scores were extremely strongly correlated (r 0.98 to 0.99) with mean scores in all BDI versions (Figures 3 and S2). Screening assignments were also

TABLE 3. Version membership and standardized one-factor model parameters of BDI-IA items

Item	Content	Item included in version			Factor loading	Item thresholds		
		BDI-SF	BDI-PC	BDI-6		1	2	3
1	Mood	✓	✓	✓	0.80	-0.51	0.64	1.52
2	Pessimism	✓	✓	–	0.75	-0.83	0.25	1.12
3	Sense of failure	✓	✓	–	0.84	-0.74	0.45	1.11
4	Lack of satisfaction	✓	✓	–	0.83	-0.55	0.44	1.46
5	Guilty feeling	✓	–	✓	0.88	-0.46	0.28	1.00
6	Sense of punishment	–	–	–	0.65	0.07	0.74	1.27
7	Negative self-view	✓	✓	–	0.84	-0.69	0.32	0.96
8	Self-accusations	–	✓	–	0.81	-0.90	-0.21	0.31
9	Self-destructiveness	✓	✓	–	0.70	-0.68	1.12	2.11
10	Crying	–	–	–	0.71	-0.48	0.26	0.69
11	Irritability	–	–	✓	0.66	-0.92	0.32	1.31
12	Social withdrawal	✓	–	–	0.74	-0.36	0.64	1.82
13	Indecisiveness	✓	–	✓	0.75	-0.68	0.05	1.60
14	Negative body image	✓	–	–	0.73	-0.13	0.36	0.63
15	Work inhibition	✓	–	✓	0.71	-0.91	-0.03	1.49
16	Sleep disturbance	–	–	–	0.56	-0.67	0.81	1.34
17	Fatigability	✓	–	✓	0.78	-0.78	0.08	1.26
18	Loss of appetite	✓	–	–	0.58	-0.05	0.73	1.54
19	Weight loss	–	–	–	0.24	0.68	1.19	1.73
20	Somatic preoccupation	–	–	–	0.42	0.25	1.48	1.93
21	Loss of libido	–	–	–	0.58	0.20	0.93	1.62

MDD = Major Depressive Disorder; BDI-IA = Beck Depression Inventory (revised version); BDI-SF = Beck Depression Inventory – Short Form; BDI-PC = Beck Depression Inventory for Primary Care; BDI-6 = Beck Depression Inventory-6.

highly similar (r 0.96 to 0.98 in the primary condition), with only 1–1.5% of respondents classified differently with the two methods (Table S8). Factor scores did not differ in discriminability from the BDI-IA sum/mean scores.

Optimal cut-offs

The BDI cut-offs optimized for agreement with diagnoses (χ) in the whole sample were slightly higher than the primary a priori value and approximately corresponded for the full BDI-IA to a sum score of 19 or greater (Figure 2a). This difference was mostly due to females, as the value was 17½ when estimated in the male subsample (Figure 2b). When the optimization criterion was an equal misclassification cost for false positives and negatives, the BDI-IA sum score equivalent was 18, and when false negatives were deemed three times as costly, this cut-off was approximately 13 (Figure S3). As with matched cut-offs, optimized cut-offs were higher for the abbreviated versions due to the higher scores for the included items than the excluded ones (Table 6).

Discussion

The present study was designed to compare the diagnostic accuracy of the BDI and its abbreviated forms in an adolescent clinical population. We found the BDI-IA to be acceptably unidimensional in our adolescent psychiatric sample. This finding is

consistent with previous studies suggesting that for adequately capturing depressive symptoms, the BDI total score might be preferable over the dimension-specific subscales (52,53). The unidimensional factorial structure of the BDI-IA in our study makes abbreviated versions of the questionnaire meaningful, as the items measure the same latent construct, and the items included in the abbreviations also had the highest factor loadings.

All the abbreviations of the BDI-IA were determined to be as good as the full scale in detecting those adolescents with diagnosed depression, with a trend towards being even better. This might be explained by the abbreviated versions focusing on the core symptoms of depression, which are most indicative of the overall construct, as defined in diagnostic systems and shown by their factor loadings. The excluded somatic items may also be less diagnostic among adolescents.

The BDI-6 was at the top in diagnostic agreement and the diagnostic odds ratio across diagnostic groupings despite being the shortest scale. Values obtained for the one item longer BDI-PC, which also includes the suicide item, were practically as good, and both can be expected to be equally good in screening adolescents, cutting the number of items by two-thirds compared to the full BDI. Interestingly, these two abbreviations share only the first Mood item, and the BDI-6 is the only

TABLE 4. Diagnostic test results, MDD vs. no depressive disorder, mean score cut-off equivalent to BDI-1A sum ≥ 16 .

Version	Items	Mean cut-off	AP	Confusion matrix				Basic ratios			κ	Diagnostic odds ratio			Number needed to diagnose		Global test		Individual test adjusted p-values	
				TP	FP	FN	TN	Sens.	Spec.	PPV		NPV	DOR	95% CI	NND	95% CI	Stat.	Adj. p	Sens.	Spec.
BDI-1A	21	0.76	71%	350	89	23	160	0.938	0.643	0.797	0.874	0.608	27.36	[16.67, 44.89]	1.72	[1.51, 2.05]	—	—	—	—
BDI-SF	13	0.82	69%	348	84	25	165	0.933	0.663	0.806	0.868	0.620	27.34	[16.86, 44.34]	1.68	[1.48, 1.99]	2.22	0.329	—	—
BDI-PC	7	0.86	72%	354	95	19	154	0.949	0.618	0.788	0.890	0.598	30.20	[17.82, 51.19]	1.76	[1.54, 2.10]	3.12	0.211	—	—
BDI-6	6	1.00	69%	349	81	24	168	0.936	0.675	0.812	0.875	0.634	30.16	[18.45, 49.29]	1.64	[1.45, 1.93]	2.36	0.307	—	—

Mean cut-off = smallest mean score yielding a positive test; AP = apparent prevalence; TP = true positives; FP = false positives; FN = false negatives; TN = true negatives; Sens. = sensitivity; Spec. = specificity; PPV/NPV = positive/negative predictive value; κ = Cohen's kappa; Global test = joint comparison of sensitivity and specificity against the BDI-1A in this condition; Stat. = test statistic; Adj. p = adjusted p-value

TABLE 5. Diagnostic test results, MDD vs. no depressive disorder, factor score cut-off equivalent to BDI-1A sum ≥ 16 .

Version	Items	Factor score cut-off	Confusion matrix				Basic ratios		κ	Diagnostic odds ratio		Number needed to diagnose		Global test		Individual test adjusted p-values			
							Spec.	PPV		NPV	DOR	95% CI	NND	95% CI	Stat.	Adj. p	Sens.	Spec.	
			TP	FP	FN	TN													Sens.
BDI-1A	21	-0.35	70%	350	85	23	164	0.938	0.659	0.805	0.877	0.623	29.36	[17.87, 48.24]	1.68	[1.48, 1.98]	2.70	0.260	—
BDI-SF	13	-0.33	70%	350	85	23	164	0.938	0.659	0.805	0.877	0.623	29.36	[17.87, 48.24]	1.68	[1.48, 1.98]	1.34	0.512	—
BDI-PC	7	-0.32	71%	350	90	23	159	0.938	0.639	0.795	0.874	0.604	26.88	[16.39, 44.09]	1.73	[1.52, 2.06]	0.04	0.980	—
BDI-6	6	-0.35	70%	350	86	23	163	0.938	0.655	0.803	0.876	0.619	28.84	[17.56, 47.37]	1.69	[1.48, 2.00]	0.36	0.835	—

Mean cut-off = smallest mean score yielding a positive test; AP = apparent prevalence; TP = true positives; FP = false positives; FN = false negatives; TN = true negatives; Sens. = sensitivity; Spec. = specificity; PPV/NPV = positive/negative predictive value; κ = Cohen's kappa; Global test = joint comparison of sensitivity and specificity against the BDI-1A in this condition; Stat. = test statistic; Adj. p = adjusted p-value

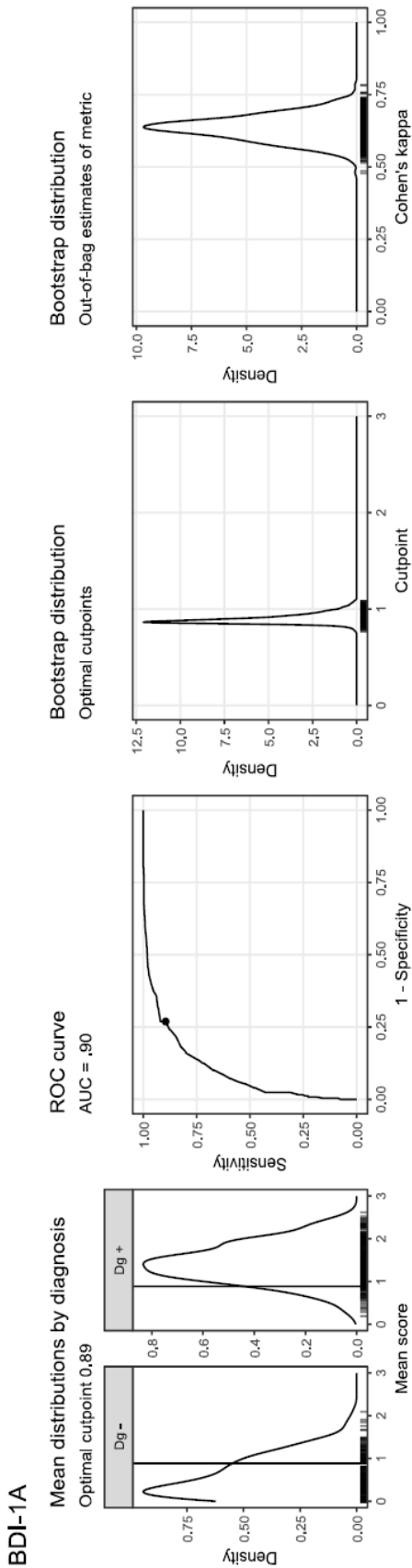


FIGURE 2a. Sample-optimal cut-offs for the main diagnostic condition, MDD vs. no depression diagnosis, criterion Cohen’s kappa, all participants together.

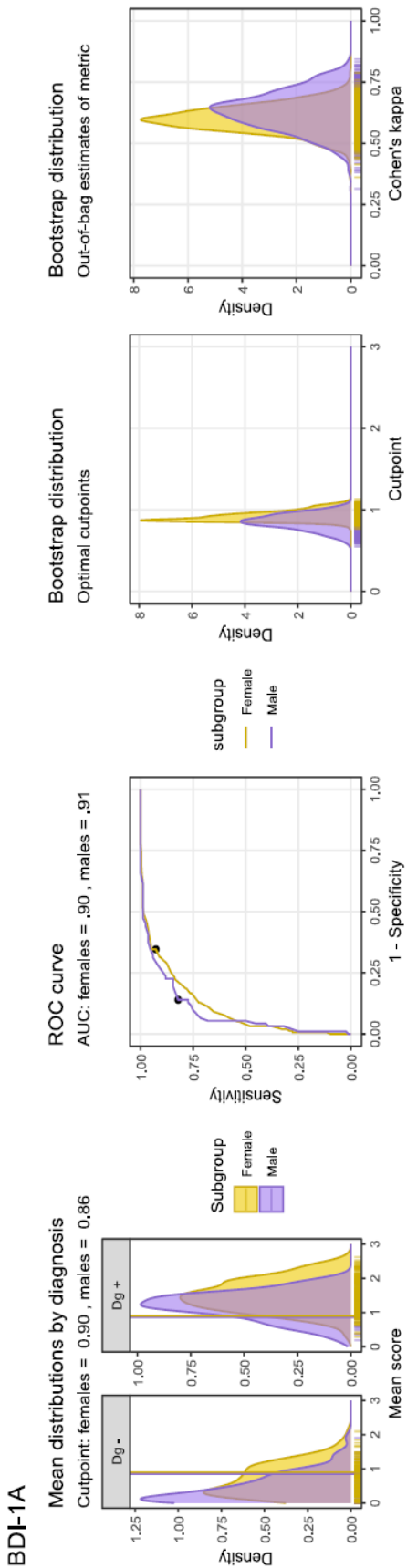


FIGURE 2b. Sample-optimal cut-offs for the main diagnostic condition, MDD vs. no depression diagnosis, criterion Cohen’s kappa, by gender.

abbreviation that includes irritability, which is a core feature of adolescent depression, as opposed to adult depression (54). In accordance with our findings, two previous studies have revealed that the BDI-PC could accurately detect MDD in pediatric care, yielding relatively high sensitivity and specificity rates (91%) (35), and a sensitivity of 81% and specificity of 90% (55).

The sensitivity analyses using two alternative categorizations of the diagnoses, which included dysthymias and assigned them to either depressed or

non-depressed groups, produced practically identical results regarding relative diagnostic accuracy. Values for the various indices were necessarily slightly poorer, as these diagnostically intermediate cases increased overlap in both diagnostic categories and BDI scores.

Using factor scores did not improve detection, despite a sample-optimized measurement model. This may, for example, be due to rarer symptoms not being indicative of greater depression, as assumed by the model with item-specific thresholds. More precise reasons need to be explored in a separate study, along with other possible alternatives to the sum score.

The sample-optimal cut-off for the full BDI was 19, which is slightly higher than the value suggested by the previous literature. Setting screening cut-offs is always a balance between sensitivity and specificity; there is no objective method to optimize discriminability without assigning a relative cost to misses and false alarms. The DOR, NND, and unweighted kappa measures used in the present study all implicitly assign such a cost. With our misclassification cost analyses, we demonstrated one way of adjusting for unequal consequences, but determining the appropriate weights, for instance, to minimize the societal burden or treatment resource use is outside the scope of our paper.

Strengths and limitations

Our study had several strengths. The sample was representative of patients in the adolescent patient population, thus enabling the generalizability of the

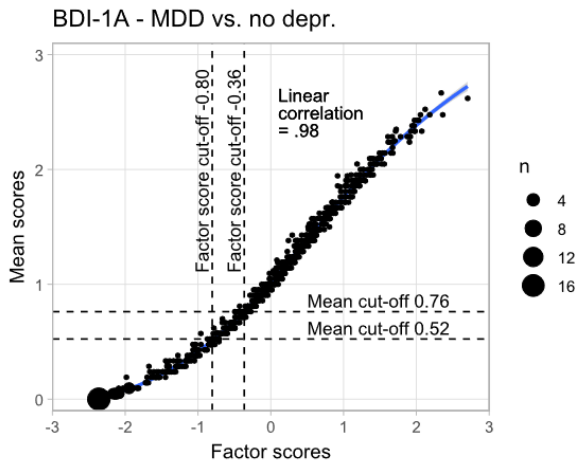


Figure 3. BDI mean scores versus factor scores with fit lines. Note: The higher and lower horizontal cut-off lines correspond to BDI-1A sums 16 and 11, respectively, and the factor score cut-offs to the matched sensitivities.

TABLE 6. Optimal cut-off results.

BDI version	Optimization criterion	Subgroup	Optimal cut-off	κ	Sensitivity	Specificity	AUC	DOR	NND
BDI-1A	Cohen's kappa	-	0.89	0.638	0.895	0.731	0.902	—	1.60
BDI-SF	Cohen's kappa	-	1.01	0.619	0.845	0.775	0.903	—	1.61
BDI-PC	Cohen's kappa	-	1.10	0.634	0.861	0.771	0.893	—	1.58
BDI-6	Cohen's kappa	-	1.04	0.615	0.866	0.743	0.900	—	1.64
BDI-1A	Cohen's kappa	Male	0.86	0.682	0.821	0.860	0.910	—	1.47
BDI-1A	Cohen's kappa	Female	0.90	0.611	0.927	0.654	0.896	—	1.72
BDI-SF	Cohen's kappa	Male	0.79	0.629	0.869	0.763	0.912	—	1.58
BDI-SF	Cohen's kappa	Female	1.07	0.597	0.886	0.699	0.896	—	1.71
BDI-PC	Cohen's kappa	Male	1.01	0.646	0.738	0.903	0.893	—	1.56
BDI-PC	Cohen's kappa	Female	1.12	0.605	0.896	0.692	0.890	—	1.70
BDI-6	Cohen's kappa	Male	0.95	0.650	0.857	0.796	0.895	—	1.53
BDI-6	Cohen's kappa	Female	1.14	0.623	0.910	0.692	0.899	—	1.66
BDI-1A	Misclassification cost 1:1	-	0.86	0.638	0.895	0.731	0.902	—	1.60
BDI-SF	Misclassification cost 1:1	-	0.89	0.606	0.893	0.699	0.903	—	1.69
BDI-PC	Misclassification cost 1:1	-	1.05	0.634	0.861	0.771	0.893	—	1.58
BDI-6	Misclassification cost 1:1	-	1.02	0.615	0.866	0.743	0.900	—	1.64
BDI-1A	Misclassification cost 1:3	-	0.62	0.599	0.962	0.602	0.902	—	1.77
BDI-SF	Misclassification cost 1:3	-	0.68	0.611	0.979	0.594	0.903	—	1.75
BDI-PC	Misclassification cost 1:3	-	0.72	0.594	0.949	0.614	0.893	—	1.77
BDI-6	Misclassification cost 1:3	-	0.74	0.565	0.957	0.574	0.900	—	1.88

κ = Cohen's kappa; AUC = Area under the curve; DOR = Diagnostic odds ratio; NND = Number needed to diagnose

findings to adolescent outpatient psychiatric care. Sensitivity and specificity are independent of prevalence, and the high values suggest that the abbreviated questionnaires are also suitable for screening in primary care. The use of gold-standard research diagnoses over register diagnoses maximized the reliability of the screening reference. In addition, our diagnostically naturalistic and heterogeneous sample was relatively large, making it sufficient in statistical power. Moreover, a paired comparative diagnostic accuracy design was used, providing greater statistical power than unpaired designs.

Regarding the limitations of this study, although SCID is widely used in clinical settings, including in Finland, and a portion of our sample consisted of older adolescents (i.e., ages 17-20), we acknowledge that this instrument may not adequately capture the developmental nuances, particularly in younger adolescents. Furthermore, the BDI-IA, being an older version, may have limitations, such as failing to differentiate between increases and decreases in depression-related vegetative symptoms (i.e., sleep and appetite) (14), despite its high correlation with the BDI-II (15). However, recent psychometric studies continue to support the utility of older versions of the BDI (56,57).

Additionally, our results naturally depend on the employed definition of depression and its operationalization as diagnostic criteria. However, there is an ongoing debate on these definitions, etiology, and the relevant pathological period. Thus, if the diagnosis had a greater emphasis on somatic symptoms, the full BDI might prove superior. Moreover, according to Finnish treatment guidelines, people with mild depression should be treated in primary health care. However, there were a few participants with mild depression in our sample who might have been referred to the outpatient clinic not primarily because of their depression but due to the presence of other comorbid conditions. Finally, although sensitivity and specificity are, in principle, independent of prevalence, our findings cannot necessarily be generalized to primary health care or general population samples. Future studies replicating our results in primary healthcare settings are therefore recommended.

Conclusion

In our study, the abbreviations of the BDI-IA proved equally effective as the full scale in detecting depressive symptoms among adolescent psychiatric outpatients. These results support earlier findings on the applicability of brief and user-friendly questionnaires to ensure optimal depression screening and minimize the administrative burden, especially in primary care settings where clinical

decision making, and appropriate referrals often need to take place within a limited time frame. Since we examined the abbreviated form items embedded in the full version of BDI-IA, additional research is required to assess the clinical utility and discriminative capacity of the abbreviated forms when used as standalone questionnaires in clinical settings. In future investigations, the BDI-PC and BDI-6 as optimal abbreviated forms of the BDI should be compared with the other brief depression screening tools in adolescent psychiatry.

Ethical considerations

The study complied with the ethical principles set by the 7th revision of the Declaration of Helsinki. The Research Ethics Committee of Kuopio University Hospital approved the study procedures (238/2017).

Funding

Researcher Tommi Tolmunen was supported by the Strategic Research Council within the Academy of Finland (SchoolWell, grant number 352509, work package 352511).

Conflict of interests

The authors declare that they have no conflict of interest.

References

1. Clayborne ZM, Varin M, Colman I. Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. *J Am Acad Child Adolesc Psychiatry*. 2019 Jan;58(1):72–9.
2. Rohde P, Lewinsohn PM, Klein DN, Seeley JR, Gau JM. Key Characteristics of Major Depressive Disorder Occurring in Childhood, Adolescence, Emerging Adulthood, and Adulthood. *Clin Psychol Sci*. 2013 Jan;1(1):41–53.
3. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global Prevalence of Depressive and Anxiety Symptoms in Children and Adolescents During COVID-19: A Meta-analysis. *JAMA Pediatr*. 2021 Nov 1;175(11):1142.
4. WHO. Finnish girls' mental health deteriorated during COVID-19 pandemic. [Internet]. 2023 [cited 2023 Sep 3]. Available from: <https://www.who.int/europe/news/item/09-03-2023-finnish-girls-mental-health-deteriorated-during-covid-19-pandemic-new-data-show>
5. Davis M, Hoskins K, Phan M, Hoffacker C, Reilly M, Fugo PB, et al. Screening Adolescents for Sensitive Health Topics in Primary Care: A Scoping Review. *J Adolesc Health*. 2022 May;70(5):706–13.
6. Beck AT. An Inventory for Measuring Depression. *Arch Gen Psychiatry*. 1961 Jun 1;4(6):561.
7. Zhang X, Savalei V. Improving the Factor Structure of Psychological Scales: The Expanded Format as an Alternative to the Likert Scale Format. *Educ Psychol Meas*. 2016 Jun;76(3):357–86.
8. Beck AT. Cognitive therapy of depression. Guilford press; 1979.
9. Beck AT. Manual for the beck depression inventory-II. San Antonio, TX: Psychological Corporation; 1996.

10. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders 4th ed. (DSM-IV). Washington, DC; 1994.
11. Basker M, Moses PD, Russell S, Russell PSS. The psychometric properties of Beck Depression Inventory for adolescent depression in a primary-care paediatric setting in India. *Child Adolesc Psychiatry Ment Health*. 2007 Dec;1(1):8.
12. Roivainen E. Beckin depressioasteikon tulkinta. *Duodecim Lääketieteellinen Aikakauskirja*. 2008 Nov 10;124:2467–70.
13. Beck AT, Steer RA. Internal consistencies of the original and revised beck depression inventory. *J Clin Psychol*. 1984 Nov;40(6):1365–7.
14. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *J Pers Assess*. 1996 Dec;67(3):588–97.
15. Dozois DJA, Dobson KS, Ahnberg JL. A psychometric evaluation of the Beck Depression Inventory–II. *Psychol Assess*. 1998 Jun;10(2):83–9.
16. Wang YP, Gorenstein C. Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Rev Bras Psiquiatr*. 2013 Dec;35(4):416–31.
17. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol*. 2006 Jan;62(1):123–46.
18. Nuevo R, Lehtinen V, Reyna-Liberato PM, Ayuso-Mateos JL. Usefulness of the Beck Depression Inventory as a screening method for depression among the general population of Finland. *Scand J Public Health*. 2009 Jan;37(1):28–34.
19. Viinamäki H, Tanskanen A, Honkalampi K, Koivumaa-Honkanen H, Haatainen K, Kaustio O, et al. Is the beck depression inventory suitable for screening major depression in different phases of the disease? *Nord J Psychiatry*. 2004 Jan;58(1):49–53.
20. Stockings E, Degenhardt L, Lee YY, Mihalopoulos C, Liu A, Hobbs M, et al. Symptom screening scales for detecting major depressive disorder in children and adolescents: A systematic review and meta-analysis of reliability, validity and diagnostic utility. *J Affect Disord*. 2015 Mar;174:447–63.
21. Wang YP, Gorenstein C. The Beck depression inventory: Uses and applications. In: *The Neuroscience of Depression* [Internet]. Elsevier; 2021 [cited 2023 Dec 27]. p. 165–74. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780128179338000207>
22. Beck AT, Beck RW. Screening Depressed Patients in Family Practice: A Rapid Technic. *Postgrad Med*. 1972 Dec;52(6):81–5.
23. Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behav Res Ther*. 1997 Aug;35(8):785–91.
24. Bech P, Gormsen L, Loldrup D, Lunde M. The clinical effect of clomipramine in chronic idiopathic pain disorder revisited using the Spielberger State Anxiety Symptom Scale (SSASS) as outcome scale. *J Affect Disord*. 2009 Dec;119(1–3):43–51.
25. Blom EH, Bech P, Högberg G, Larsson JO, Serlachius E. Screening for depressed mood in an adolescent psychiatric context by brief self-assessment scales – testing psychometric validity of WHO-5 and BDI-6 indices by latent trait analyses. *Health Qual Life Outcomes*. 2012;10(1):149.
26. Lee A, Park J. Diagnostic Test Accuracy of the Beck Depression Inventory for Detecting Major Depression in Adolescents: A Systematic Review and Meta-Analysis. *Clin Nurs Res*. 2022 Nov;31(8):1481–90.
27. Ambrosini PJ, Metz C, Bianchi MD, Rabinovich H, Undie A. Concurrent Validity and Psychometric Properties of the Beck Depression Inventory in Outpatient Adolescents. *J Am Acad Child Adolesc Psychiatry*. 1991 Jan;30(1):51–7.
28. Blom EH, Larsson JO, Serlachius E, Ingvar M. The differentiation between depressive and anxious adolescent females and controls by behavioural self-rating scales. *J Affect Disord*. 2010 May;122(3):232–40.
29. Dolle K, Schulte-Körne G, O’Leary AM, Von Hofacker N, Izat Y, Allgaier AK. The Beck Depression Inventory-II in adolescent mental health patients: Cut-off scores for detecting depression and rating severity. *Psychiatry Res*. 2012 Dec;200(2–3):843–8.
30. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191–4.
31. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig I, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015 Oct 28;351:h5527.
32. First MB, Spitzer RL, Gibbon M, Williams, JBW. Structured clinical interview for DSM-IV axis I disorders, clinician version (SCID-CV). Washington, DC: American Psychiatric Press; 1996. 132 p.
33. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev*. 1988 Jan;8(1):77–100.
34. Steer RA, Cavalieri TA, Leonard DM, Beck AT. Use of the Beck depression inventory for primary care to screen for major depression disorders. *Gen Hosp Psychiatry*. 1999 Mar;21(2):106–11.
35. Winter LB, Steer RA, Jones-Hicks L, Beck AT. Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory for Primary Care. *J Adolesc Health*. 1999 Jun;24(6):389–94.
36. Aalto AM, Elovainio M, Kivimäki M, Uutela A, Pirkola S. The Beck Depression Inventory and General Health Questionnaire as measures of depression in the general population: A validation study using the Composite International Diagnostic Interview as the gold standard. *Psychiatry Res*. 2012 May;197(1–2):163–71.
37. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available from: <http://www.r-project.org>
38. Neubert K, Brunner E. A studentized permutation test for the non-parametric Behrens–Fisher problem. *Comput Stat Data Anal*. 2007 Jun;51(10):5192–204.
39. Karch JD. Psychologists Should Use Brunner-Munzel’s Instead of Mann-Whitney’s U Test as the Default Nonparametric Procedure. *Adv Methods Pract Psychol Sci*. 2021 Apr;4(2):251524592199960.
40. Ara T. Brunnermunzel: (Permuted) Brunner-Munzel Test. [Internet]. 2022. Available from: <https://CRAN.R-project.org/package=brunnermunzel>
41. Akoglu H. User’s guide to sample size estimation in diagnostic accuracy studies. *Turk J Emerg Med*. 2022;22(4):177.
42. Noether GE. Sample Size Determination for Some Common Nonparametric Tests. *J Am Stat Assoc*. 1987 Jun;82(398):645–7.
43. Konietzschke F, Brunner E. rankFD: An R Software Package for Nonparametric Analysis of General Factorial Designs. *R J*. 2023 Aug 26;15(1):142–58.

44. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw.* 2012 May;48:1–36.
45. Stevenson M, Sergeant E. epiR: Tools for the Analysis of Epidemiological Data. [Internet]. 2024 [cited 2024 Jul 1]. Available from: <https://CRAN.R-project.org/package=epiR>
46. Linn S, Grunau PD. New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests. *Epidemiol Perspect Innov.* 2006 Dec;3(1):11.
47. Roldán-Nofuentes JA. Compbdt: an R program to compare two binary diagnostic tests subject to a paired design. *BMC Med Res Methodol.* 2020 Dec;20(1):143.
48. Wilson KJ, Henrion M, Roldán Nofuentes JA. testCompareR: Comparing Two Diagnostic Tests with Dichotomous Results using Paired Data. [Internet]. 2024 [cited 2024 Jul 3]. Available from: <https://CRAN.R-project.org/package=testCompareR>
49. Thiele C, Hirschfeld G. cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. *J Stat Softw.* 2021 Jun;98:1–27.
50. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016 Nov;6(11):e012799.
51. Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. *BMJ.* 2012 Jul 2;345(jul02 1):e3999–e3999.
52. Araya R, Montero-Marín J, Barroilhet S, Fritsch R, Montgomery A. Detecting depression among adolescents in Santiago, Chile: sex differences. *BMC Psychiatry.* 2013 Dec;13(1):122.
53. Keller F, Kirschbaum-Lesch I, Straub J. Factor Structure and Measurement Invariance Across Gender of the Beck Depression Inventory-II in Adolescent Psychiatric Patients. *Front Psychiatry.* 2020 Dec 23;11:527559.
54. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders [Internet]. Fifth Edition. American Psychiatric Association; 2013 [cited 2024 Mar 30]. Available from: <https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596>
55. Pietsch K, Hoyler A, Frühe B, Kruse J, Schulte-Körne G, Allgaier AK. Früherkennung von Depressionen in der Pädiatrie: Kriteriumsvalidität des Beck Depressions-Inventar Revison (BDI-II) und des Beck Depressions-Inventar–Fast Screen for Medical Patients (BDI-FS). *PPmP - Psychother · Psychosom · Med Psychol.* 2012 Jun 21;62(11):418–24.
56. D'Iorio A, Maggi G, Guida P, Aiello EN, Poletti B, Silani V, et al. Early Detection of Depression in Parkinson's Disease: Psychometrics and Diagnostics of the Spanish Version of the Beck Depression Inventory. *Arch Clin Neuropsychol.* 2024 May 21;39(4):418–22.
57. Rodríguez-Pérez V, Piñeirua Menéndez A, Ramírez-Rentería C, Mata Marín JA. Beck Depression Inventory (BDI-IA) adapted for HIV: Psychometric properties, sensitivity & specificity in depressive episodes, adjustment disorder & without symptomatology. *Salud Ment.* 2021 Dec 2;44(6):287–94.