Use of the PROMIS Depression Scale and the Beck Depression Inventory in Patients with Heart Failure

Kenneth E. Freedland, Ph.D.,
Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

Brian C. Steinmeyer, M.S.,
Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

Robert M. Carney, Ph.D.,
Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

Eugene H. Rubin, M.D., Ph.D., and
Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

Michael W. Rich, M.D.
Department of Medicine, Washington University School of Medicine, St. Louis, Missouri

Abstract

Objective: This study evaluated agreement between the PROMIS Depression scale and the Beck Depression Inventory (BDI-II) in patients with heart failure and comorbid major depression.

Method: The BDI-II and the computerized adaptive test version of the PROMIS Depression scale were administered at baseline to 158 participants in a randomized controlled trial of cognitive behavior therapy for major depression in patients with heart failure. A crosswalk table (Choi, Schalet, Cook, & Cella, 2014) was used to transform the PROMIS scores into "linked" BDI-II equivalent scores. Bland-Altman plots, histograms, and scatterplots were used to visualize the agreement between these scores at baseline and 6 months, and intraclass correlation coefficients (ICCs) were calculated for each occasion to quantify the agreement. Treatment effects and change scores were also examined.

Results: The measures agreed moderately at baseline (ICC, 0.52; p<.0001) and strongly at 6 months (ICC, 0.77; p<.0001), but on average, the linked and observed BDI-II scores differed by 3.1 points at baseline (p<.0001) and −0.17 points at 6 months (p=.78). The discrepancies were considerably larger in many individual cases on both occasions.

Conclusions: The PROMIS Depression scale is likely to play an important role in research on depression in patients with heart failure, but for now, it should be used in addition to rather than instead of the BDI-II in studies in which the BDI-II would ordinarily be used. Additional research is needed to evaluate the validity and utility of the PROMIS Depression scale in patients with heart failure.

Correspondence concerning this article should be addressed to Kenneth E. Freedland, Ph.D., Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63108. freedlak@wustl.edu.

The role and treatment of depression in patients with heart disease has been studied extensively over the past several decades. The 21-item Beck Depression Inventory (BDI-II) (Beck, Steer, & Brown, 1996) is the most widely used measure of depression in research on cardiac patients. It has been used in numerous observational, epidemiological, and treatment studies (e.g., Blumenthal et al., 2012; Carney, Freedland, Steinmeyer, Rubin, & Rich, 2016; Sherwood et al., 2011; Wei et al., 2014).

The PROMIS Depression Scale (Pilkonis et al., 2011) is newer than the BDI-II, and state-of-the-art psychometric methods were used to develop it. Although a standard short form is available, the PROMIS Depression item bank can be used to construct customized short forms for specific research or clinical applications (Cella, Gershon, Lai, & Choi, 2007). In addition, the PROMIS computerized adaptive test (CAT) application (Gershon, Rothrock, Hanrahan, Bass, & Cella, 2010) makes it possible to obtain a PROMIS Depression score after administering relatively few items. The more items that are administered, the higher the score correlates with the score obtained from the full item bank. However, the correlation exceeds 0.95 when as few as 5 items are administered (Choi, Reise, Pilkonis, Hays, & Cella, 2010). Since lengthy questionnaires can be burdensome for medically ill patients, this makes the CAT version of the PROMIS Depression scale especially appealing for research on medical patient populations. Also, data on the Depression scale and other PROMIS measures are being collected in many different patient populations, and they are being used in numerous clinical trials. Consequently, use of the PROMIS Depression scale will facilitate future comparisons of patient-reported outcomes across populations and interventions. So far, however, the PROMIS Depression scale has not been used in very many studies of cardiac patients in general or in any diagnostic subgroups of cardiac patients. To our knowledge, it has been used to date in only four studies of patients with heart failure (Fischer et al., 2014; Flynn et al., 2015; Freedland, Carney, Rich, Steinmeyer, & Rubin, 2015; Schalet et al., 2016).

Freedland et al. (2015) used the PROMIS Depression scale as a secondary outcome measure in a randomized controlled trial of cognitive behavior therapy for major depression and inadequate self-care in 158 patients with heart failure. The primary outcome measure was the BDI-II at 6 months, and the PROMIS Depression score at 6 months was a secondary outcome. The results on both measures supported the conclusion that CBT was superior to usual care for major depression in patients with heart failure.

However, the fact that similar results were obtained on these two measures does not necessarily mean that they are interchangeable. When there are two different ways to measure the same construct, it is useful to determine the extent to which the measures agree with one another. If there is high agreement between two different measures of depression, most individuals will appear to be equally severely depressed on both measures. In contrast, if there is low agreement, an individual might register as considerably more (or less) severely depressed on one measure than on the other. The present study uses baseline and 6-

month (post-treatment) data from the Freedland et al. (2015) trial to evaluate the agreement between the BDI-II and the PROMIS Depression scale in patients with heart failure and comorbid major depression.

Method

Participants

The participants (n = 158) had been diagnosed with New York Heart Association (NYHA) Class I, II, or III heart failure at least three months before enrollment. To be enrolled in the trial, they also had to meet the DSM-IV criteria for major depression and score $\geq$14 on the BDI-II. The exclusion criteria were (1) inability to participate due to cognitive impairment, frailty, a communication deficit, or a logistical barrier; (2) poor 1-year prognosis due to a noncardiac comorbidity; (3) hospitalization within the past month; (4) suicidality, psychosis, or substance abuse; or (5) initiation of an antidepressant within the past 8 weeks. The average age of the participants was 55.8 ± 11.2 years, 46.2% were women, 63.3% were white, 46.8% were married, 85.4% had at least 12 years of education, and 21.5% were employed. At enrollment, 57.6% of the participants were in NYHA Class I or II and 42.4% were in Class III. Of the 158 participants, 123 (78%) completed the BDI-II and PROMIS Depression measure at 6 months. The study was approved by the institutional review board at Washington University School of Medicine in St. Louis, and all participants provided written informed consent.

Procedure

All baseline measures were obtained immediately prior to randomization, and the 6-month measures were obtained within one week after the end of the intervention phase. Participants completed the assessments in a testing room at our clinical research center. The BDI-II was administered as a paper-and-pencil questionnaire. The web-based PROMIS Assessment Center application was used to administer the PROMIS Depression scale as a computerized adaptive test. The Assessment Center application administers one item at a time; presentation of each item (after the first one) is determined by the participant's response to the previous item. The test length averaged 5.1 ± 0.6 items per participant. Further information about the study procedures and primary results are provided in Freedland et al. (2015).

Choi and colleagues (Choi et al., 2014) produced cross-walk tables linking several popular depression instruments to PROMIS Depression scores. A representative panel of 1,120 respondents from the general population provided the data for the BDI-II analyses. Each of the 21 items on the BDI-II is rated on a 0–3 scale; consequently, BDI-II total scores can range from 0 to 63. The mean total BDI-II score in their sample was 13.7 ± 12.2, just under the $\geq$14 cutoff score for depression on the BDI-II. Two hundred eighty-nine (28%) of the participants scored $\geq$20, consistent with moderate depression. PROMIS measures are reported as *T* scores, with a mean of 50 and a standard deviation of 10 in the general population. Choi et al.'s results show that a BDI-II score of 14 is equivalent to a score of 55.6 on the PROMIS scale in the general population. The BDI-II cutoff score for moderate depression is 20; this is equivalent to a score of 59.3 on the PROMIS Depression scale. The

BDI-II cutoff score for severe depression is 29, equivalent to 64.3 on the PROMIS Depression scale.

The present study compared each participant's observed baseline and 6-month BDI-II scores to his or her BDI-II linked score based on his or her PROMIS Depression score. The linked scores were determined by transforming the individual's PROMIS Depression score into a BDI-II equivalent (linked) score from the table included in the Choi et al. (2014) report. Several methods were used to compare the measures at each occasion. First, scatterplots and linear regression models were used to examine the relationship between the observed and linked BDI-II scores. Second, Bland-Altman plots (Bland & Altman, 1986, 1995) were used to evaluate the agreement between the observed and linked BDI-II scores. These plots display the average of the two scores on the X axis, and the difference between them on the Y axis. Thus, the Bland-Altman plots display the relationship between the severity of depression and the difference between the measures within individuals. Spearman correlations were used to estimate the strength of the linear association between the severity and difference dimensions. The bias line on each plot represents the mean difference between the scores, across all levels of severity. Paired t-tests were used to determine whether the bias was significantly different from zero. The plots also display the upper (U) and lower (L) limits of agreement (LOA), within which 95% of the within-individual differences between the measures are expected to fall. Histograms were also produced to complement the Bland-Altman plots. The histograms display the frequency with which the scores either were identical or differed to varying degrees within individuals, across all levels of depression severity. In addition, intraclass correlation coefficients (ICCs) were computed for each occasion to quantify the overall level of agreement. The ICC values were interpreted according to the ranges recommended by Watson and Petrie (2010). A Pearson correlation coefficient was used to examine the linear relationship between observed and linked BDI-II pre-post change scores. Finally, Cohen's *d* was computed for the treatment effect on each measure, based on the covariate-adjusted analyses presented in the primary outcomes report (Freedland et al., 2015), in order to examine whether the measures may differ as to their sensitivity to treatment effects.

Results

Table 1 displays the means, standard deviations, and ranges for the PROMIS Depression, BDI-II observed scores, and BDI-II linked scores at baseline and at the 6-month post-treatment assessment. The covariate-adjusted differences between the intervention and usual care arms were −4.43 (95% C.I., −7.68 to −1.18; p=.008) on the observed BDI-II; −6.9 (95% C.I., −10.49 to -3.28; p=.0002) on the BDI-II linked scores based on the PROMIS crosswalk; and −6.49 (95% C.I., −9.14 to −3.84; p<.0001) on the PROMIS Depression T-score. The treatment effect sizes were *d* = 0.42 (95% C.I., 0.11 to 0.74) for the observed BDI-II and *d* = 0.79 (95% C.I., 0.48 to 1.13) for the PROMIS Depression score. Thus, the intervention had a considerably larger covariate-adjusted effect on the PROMIS Depression score than it had on the observed BDI-II total score.

Table 1 also displays the mean differences at baseline and 6 months between the observed and linked BDI-II scores. At baseline, the average observed BDI-II score was 3.11 points

higher than the linked BDI-II score (95% C.I., 1.64 to 4.57; t = 4.19; p < .0001). At 6 months, the average observed and linked BDI-II scores differed by only −0.17 points (95% C.I., −1.39 to 1.05; t = -0.28; p = .78). These differences are represented as the bias lines on Figures 1b and 2b.

Figure 1a displays a scatterplot and regression analysis of the relationship between the BDI-II observed and linked scores at baseline. Based on the fitted curve, cutoff scores on BDI-II linked scale are biased estimators of cutoff scores on the observed BDI-II. For example, the plot suggests that an individual who scores 14 on the linked BDI-II would be expected to score approximately 25 on the actual BDI-II. Similarly, an individual who scores 20 (the lower limit of the moderate range) on the linked BDI-II would be expected to score approximately 27 on the actual BDI-II, and one who scores 29 (the lower limit of the severe range) on the linked BDI-II would be expected to score approximately 32 on the actual BDI-II. Thus, the linked scores tend to overestimate the actual BDI-II scores, and the overestimates tend to be larger in the relatively mild range of major depression than in the moderate-to-severe range.

Figure 1b displays the Bland-Altman plot for the baseline data. The bias line indicates that means of the observed and linked BDI-II scores differ by 3.1 points. The upper and lower limits of agreement show that 95% of the within-individual differences between the observed and linked BDI-II scores at baseline are between −15.3 and 21.5 points. The Spearman correlation is -0.24 (p<.01), consistent with a moderate inverse linear relationship between the severity of depression and the difference between the observed and linked BDI-II scores. This suggests that while observed BDI-II scores tend to be higher than linked BDI-II scores in mildly depressed patients, observed scores tend to be lower than linked scores in more severely depressed patients. The overall agreement between the observed and the linked BDI-II scores was moderate (ICC, 0.52; 95% C.I., 0.38 to 0.64); p<.0001) at baseline. Figure 1c displays a histogram of the frequency of differences between the observed and linked BDI-II scores.

Figures 2a, 2b, and 2c display the scatterplot and regression analysis, the Bland-Altman plot, and the histogram for the 6-month data. Like the baseline results presented above, the 6-month regression analysis suggests that the standard cutoff scores on the linked BDI-II map onto different scores on the actual BDI-II. Unlike the baseline results, the Bland-Altman plot shows no significant bias between the average BDI-II linked score and the average BDI-II observed score. Overall, there was substantial agreement between the observed and linked BDI-II scores at 6 months (ICC, 0.77; 95% C.I., 0.69 to 0.83; p <.0001). However, the upper and lower limits of agreement remain relatively far apart (−13.7 to 13.4). The Spearman correlation is negligible, indicating that at 6 months there is no relationship between the severity of depression and the differences between the scores.

Finally, baseline to 6-month change scores were computed for the observed and linked BDI-II scores. The Pearson correlation between the change scores was r = 0.55 (p<.0001), suggesting that there is a moderately strong, positive linear relationship between changes over time in observed and linked BDI-II scores.

Discussion

Thanks to the work of Choi and colleagues (Choi et al., 2014), group means on the PROMIS Depression scale that are obtained in studies based in the general population can be translated into BDI-II score equivalents. The primary purpose of the present study was to determine whether this can be done in patients with heart failure and comorbid major depression, and whether the measures agree so well that they are essentially interchangeable. The results suggest that they agree relatively well but they are not interchangeable. On average, the estimated and observed scores differed by about 3 points at baseline, a non-trivial difference. On the other hand, there was almost no difference between the group means on the observed and linked BDI-II scores after treatment.

The contrast between the baseline and post-treatment findings raises the possibility of differential item functioning (DIF) between patients who have heart failure with comorbid major depression and those whose depression has improved or remitted. An earlier general population-based study found evidence of modest DIF by age or gender on several PROMIS Depression items (Teresi et al., 2009). Interestingly, two of the items ("I had trouble enjoying things that I used to enjoy" and "I felt I had no energy") are ones that are often endorsed by patients with chronic heart failure whether or not depression is present. Because the BDI includes nonspecific somatic symptoms such as fatigue, it is not unusual for BDI total scores to be mildly elevated in nondepressed cardiac patients. For example, the nondepressed comparison group (n=25) in a recent study of patients with chronic heart failure scored 4.74 ± 2.12 on the BDI (Xiong et al., 2015). An IRT study of the BDI in 1135 patients with a history of acute myocardial infarction (Wardenaar, Wanders, Roest, Meijer, & De Jonge, 2015) found that at low severity levels, the BDI predominantly detects somatic symptoms whereas at higher severity levels it measures mood and cognitive symptoms. These findings suggest the need for a study of DIF by severity of depression and treatment status on the PROMIS Depression scale and the BDI-II in patients with coronary heart disease or chronic heart failure.

Both at baseline and at 6 months, there were many outliers with large discrepancies between the observed and estimated scores. The observed BDI-II score was substantially higher than the linked BDI-II score in many cases, and substantially lower in many others. The availability of the crosswalk may make it tempting to translate individual PROMIS Depression scores into BDI-II score equivalents. However, the authors of the crosswalk recommend using it for group means but not for individual scores (Choi et al., 2014). The present study confirms that in research on patients with heart failure and comorbid major depression, BDI-II linked scores can substantially under- or over-estimate an individual's actual BDI-II score.

This study is limited by the fact that all participants had major depression at baseline. Studies of agreement between the PROMIS Depression scale and the BDI-II are needed in a broader range of patients with heart failure. In addition, the computerized adaptive test parameters were set to ensure that no more than about 5 or 6 PROMIS items would be administered in most cases. The precision of the PROMIS Depression score is related to the number of items administered. Consequently, there might have been better agreement (i.e.,

less bias) between the measures if the PROMIS computerized adaptive test application had been configured to administer more Depression items.

The PROMIS Depression scale has many strengths and is likely to play an increasingly important role in research on patients with heart failure. In our randomized controlled trial of cognitive behavior therapy for major depression in patients with heart failure, the intervention effect size was much larger on the PROMIS Depression scale than it was on the BDI-II. Thus, it is possible that the PROMIS Depression scale is more sensitive than the BDI-II to treatment-related changes in depression. This possibility, along with the relatively low participant burden and high acceptability of computerized adaptive testing, make the PROMIS Depression scale a strong candidate for use in future treatment trials.

The PROMIS Depression scale may be the best depression measure to use in many studies. However, the present results suggest that it is not completely interchangeable with the BDI-II. Further research is needed to evaluate the validity and utility of the PROMIS Depression scale in patients with heart failure, as well as in patients with other cardiovascular conditions. At least for now, the PROMIS Depression scale should be used in addition to, not instead of, the BDI-II in studies of cardiac patients in which the BDI-II would ordinarily be used.

Acknowledgments

References

Beck AT, Steer RA, & Brown GK (1996). BDI-II Manual San Antonio, Texax: The Psychological Corporation.

Bland JM, & Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1(8476), 307–310. [PubMed: 2868172]

Bland JM, & Altman DG (1995). Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet, 346(8982), 1085–1087. [PubMed: 7564793]

Blumenthal JA, Babyak MA, O'Connor C, Keteyian S, Landzberg J, Howlett J, … Whellan DJ (2012). Effects of exercise training on depressive symptoms in patients with chronic heart failure: the HF-ACTION randomized trial. JAMA, 308(5), 465–474. doi:10.1001/jama.2012.8720 [PubMed: 22851113]

Carney RM, Freedland KE, Steinmeyer BC, Rubin EH, & Rich MW (2016). Clinical predictors of depression treatment outcomes in patients with coronary heart disease. J Psychosom Res, 88, 36–41. doi:10.1016/j.jpsychores.2016.07.011 [PubMed: 27521651]

Cella D, Gershon R, Lai JS, & Choi S (2007). The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res, 16 Suppl 1, 133–141. doi:10.1007/s11136-007-9204-6 [PubMed: 17401637]

Choi SW, Reise SP, Pilkonis PA, Hays RD, & Cella D (2010). Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. Qual Life Res, 19(1), 125–136. doi:10.1007/s11136-009-9560-5 [PubMed: 19941077]

Choi SW, Schalet B, Cook KF, & Cella D (2014). Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol Assess, 26(2), 513–527. doi:10.1037/a0035768 [PubMed: 24548149]

Fischer HF, Klug C, Roeper K, Blozik E, Edelmann F, Eisele M, … Herrmann-Lingen C (2014). Screening for mental disorders in heart failure patients using computer-adaptive tests. Qual Life Res, 23(5), 1609–1618. doi:10.1007/s11136-013-0599-y [PubMed: 24338104]

Flynn KE, Dew MA, Lin L, Fawzy M, Graham FL, Hahn EA, … Weinfurt KP (2015). Reliability and construct validity of PROMIS(R) measures for patients with heart failure who undergo heart transplant. Qual Life Res, 24(11), 2591–2599. doi:10.1007/s11136-015-1010-y [PubMed: 26038213]

Freedland KE, Carney RM, Rich MW, Steinmeyer BC, & Rubin EH (2015). Cognitive behavior therapy for depression and self-care in heart failure patients: A randomized clinical trial. JAMA Internal Medicine, 175(11), 1773–1782. [PubMed: 26414759]

Gershon RC, Rothrock N, Hanrahan R, Bass M, & Cella D (2010). The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. J Appl Meas, 11(3), 304–314. [PubMed: 20847477]

Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, & Cella D (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. Assessment, 18(3), 263–283. doi: 10.1177/1073191111411667 [PubMed: 21697139]

Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, … Cella D (2016). Clinical validity of PROMIS Depression, Anxiety, and Anger across diverse clinical samples. J Clin Epidemiol, 73, 119–127. doi:10.1016/j.jclinepi.2015.08.036 [PubMed: 26931289]

Sherwood A, Blumenthal JA, Hinderliter AL, Koch GG, Adams KF Jr., Dupree CS, … O'Connor CM (2011). Worsening depressive symptoms are associated with adverse clinical outcomes in patients with heart failure. J Am Coll Cardiol, 57(4), 418–423. doi:10.1016/j.jacc.2010.09.031 [PubMed: 21251581]

Teresi JA, Ocepek-Welikson K, Kleinman M, Eimicke JP, Crane PK, Jones RN, … Cella D (2009). Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. Psychol Sci Q, 51(2), 148–180. [PubMed: 20336180]

Wardenaar KJ, Wanders RB, Roest AM, Meijer RR, & De Jonge P (2015). What does the beck depression inventory measure in myocardial infarction patients? a psychometric approach using item response theory and person-fit. Int J Methods Psychiatr Res, 24(2), 130–142. doi:10.1002/mpr.1467 [PubMed: 25994207]

Watson PF, & Petrie A (2010). Method agreement analysis: a review of correct methodology. Theriogenology, 73(9), 1167–1179. doi:10.1016/j.theriogenology.2010.01.003 [PubMed: 20138353]

Wei J, Pimple P, Shah AJ, Rooks C, Bremner JD, Nye JA, … Vaccarino V (2014). Depressive symptoms are associated with mental stress-induced myocardial ischemia after acute myocardial infarction. PLoS One, 9(7), e102986. doi:10.1371/journal.pone.0102986 [PubMed: 25061993]

Xiong GL, Prybol K, Boyle SH, Hall R, Streilein RD, Steffens DC, … Jiang W (2015). Inflammation Markers and Major Depressive Disorder in Patients With Chronic Heart Failure: Results From the Sertraline Against Depression and Heart Disease in Chronic Heart Failure Study. Psychosom Med, 77(7), 808–815. doi:10.1097/psy.0000000000000216 [PubMed: 26186432]
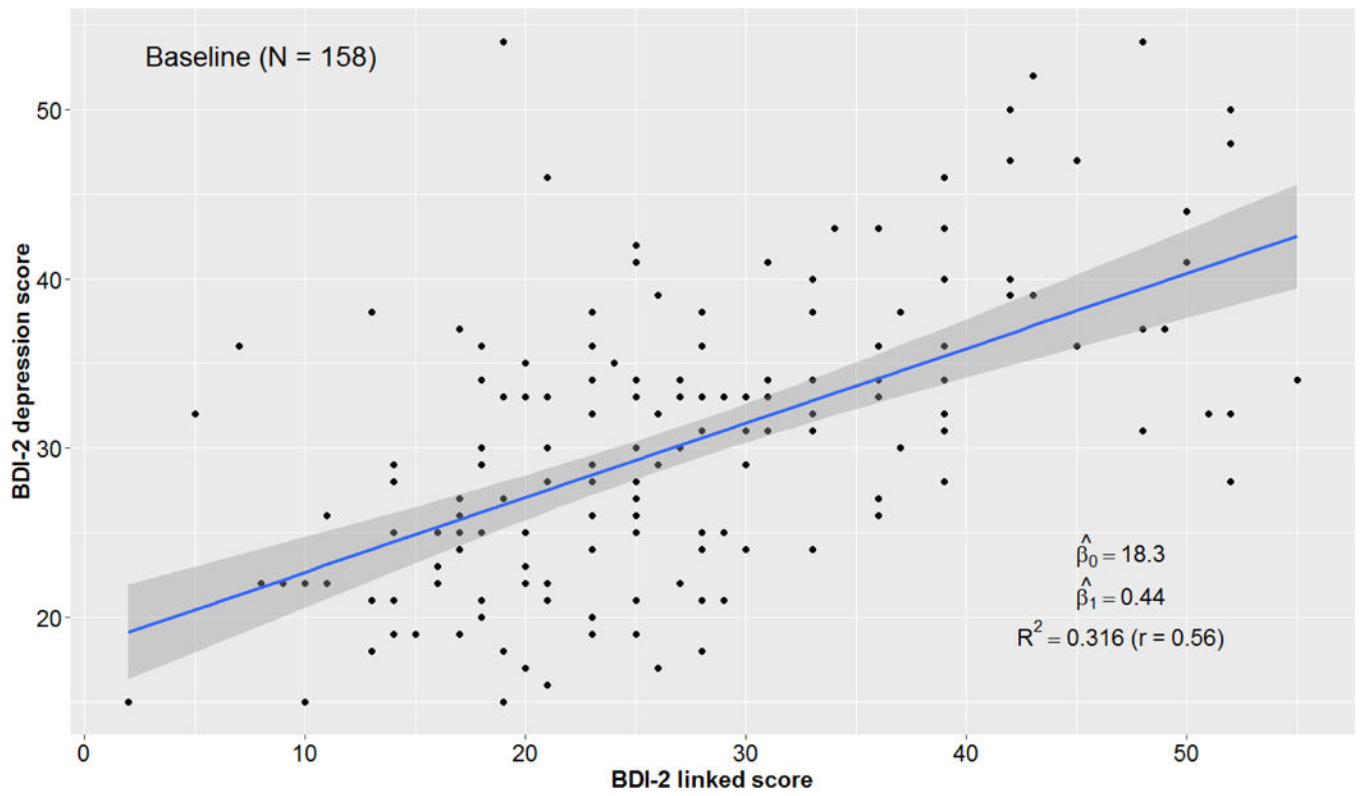
Figure 1a.

Scatterplot and regression analysis of the association of baseline BDI-II linked scores, based on PROMIS Depression scores, with baseline BDI-II observed scores. The gray zone represents the standard error of the regression line.
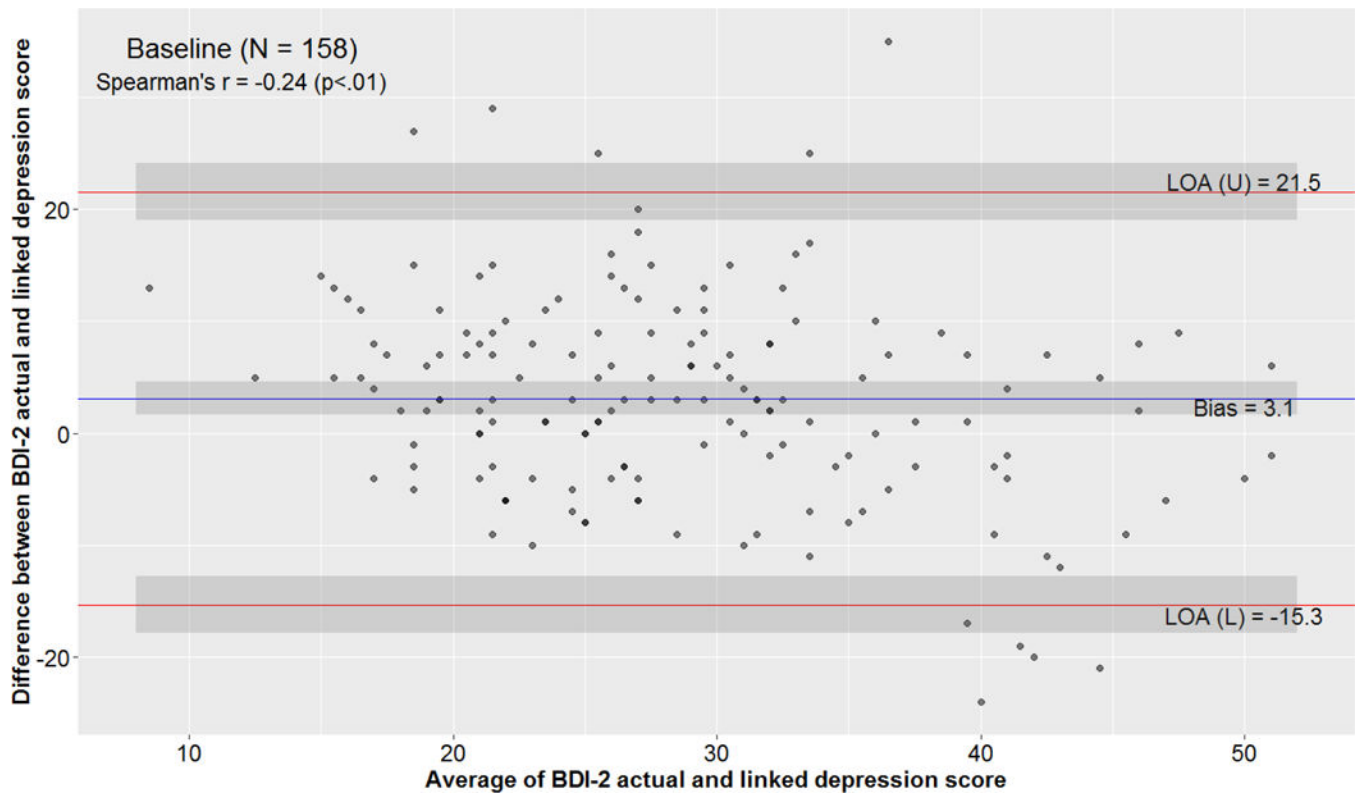
Figure 1b.

Bland-Altman plot comparing observed and linked BDI-II scores within individuals. The severity of depression is represented on the X-axis as the average of the observed and linked BDI-II scores. Agreement is represented on the Y-axis as the difference between the observed and linked BDI-II scores. The bias line indicates the mean difference between the observed and linked scores; 95% of the differences are expected to fall within the upper and lower limits of agreement (LOA). The Spearman correlation coefficient indicates an inverse linear association between the severity of depression and the difference between the observed and linked scores.
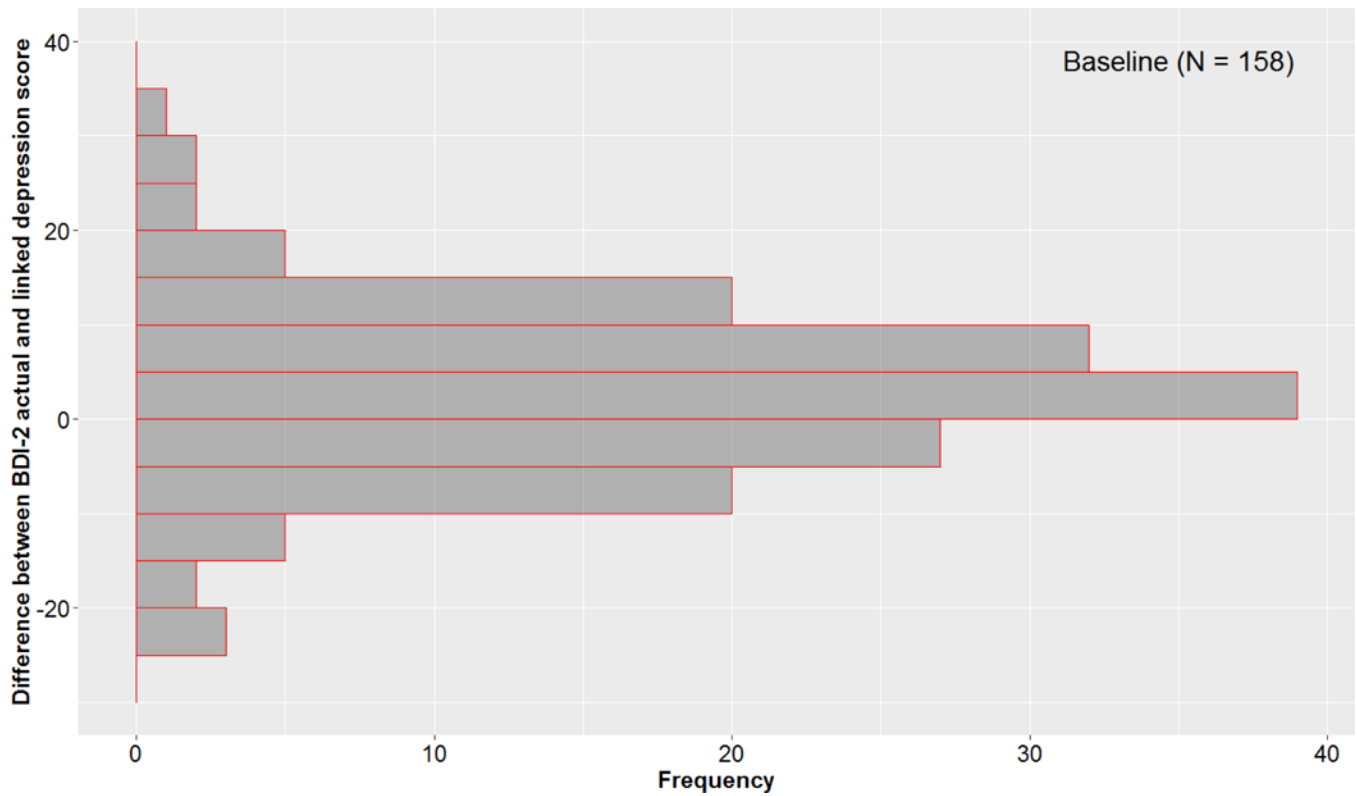
Figure 1c.
Histogram of differences between observed and linked baseline BDI-II scores, collapsed over all levels of severity. The vertical axis represents the difference between the scores, the horizontal axis represents the number of participants, and the bars depict 5-point increments along the vertical axis. The graph shows that while extreme within-participant differences between the scores were rare, small-to-moderate differences were common. It also shows that the differences were bidirectional.
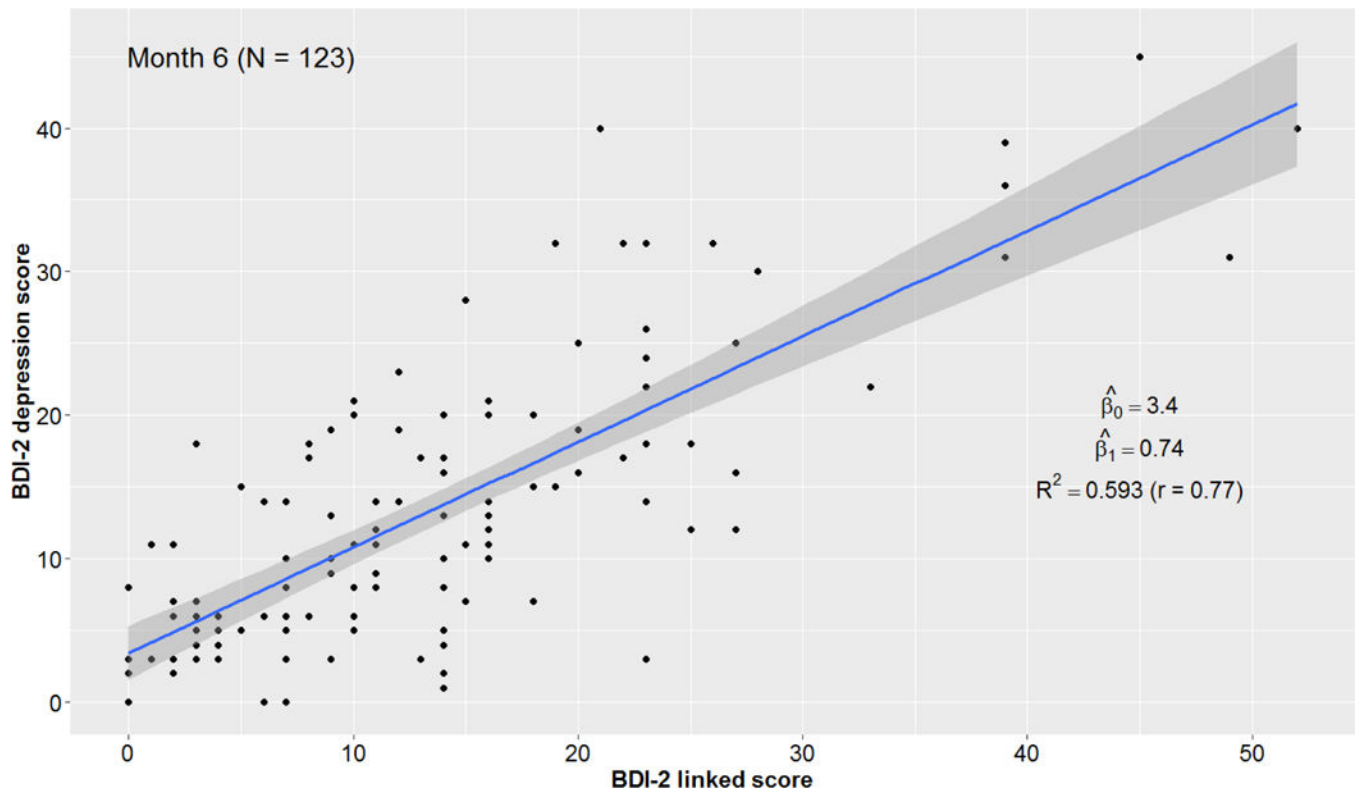
Figure 2a.
Scatterplot and regression analysis of the association of 6-month BDI-II linked scores, based on PROMIS Depression scores, with 6-month BDI-II observed scores. The gray zone represents the standard error of the regression line.
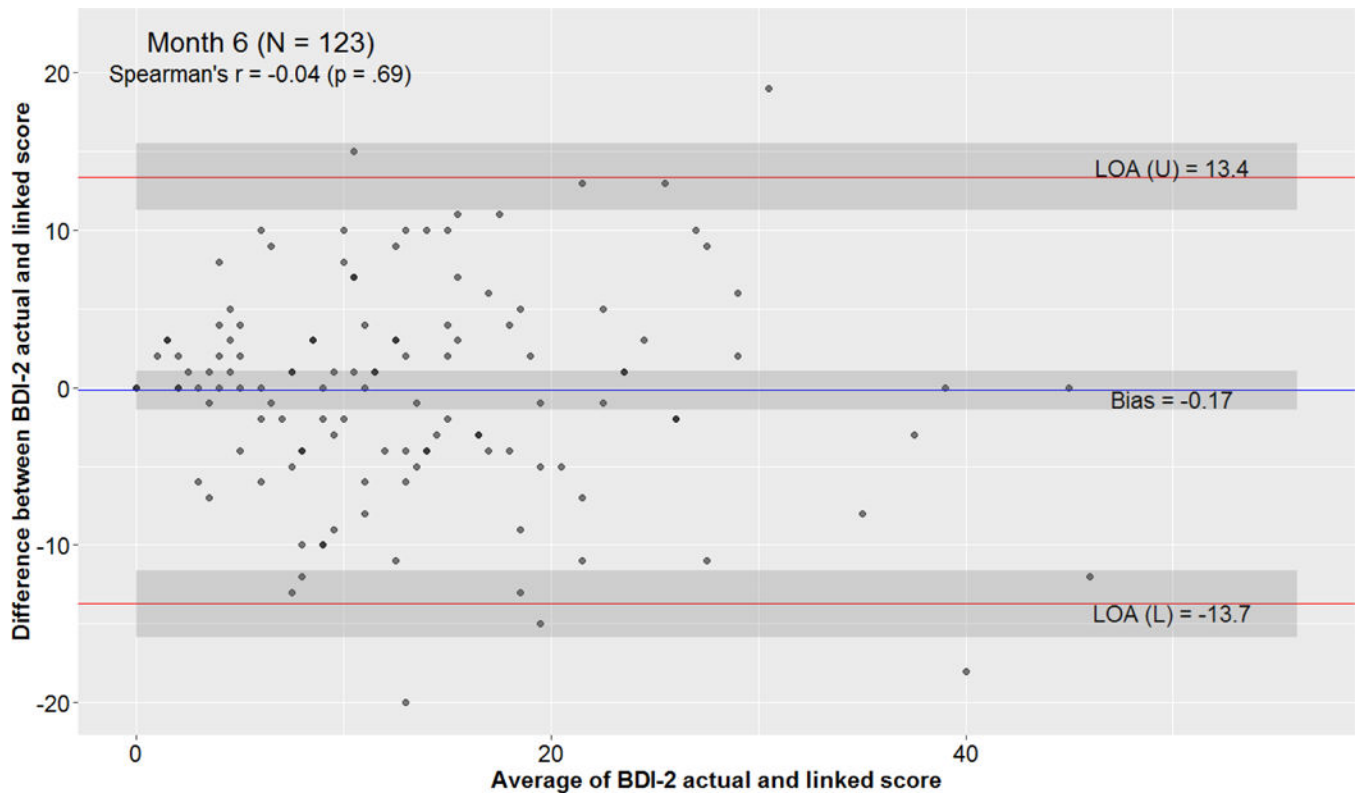
Figure 2b.

Bland-Altman plot comparing observed and linked BDI-II scores at 6 months. The severity of depression is represented on the X-axis as the average of the observed and linked BDI-II scores. Agreement is represented on the Y-axis as the difference between the observed and linked BDI-II scores. The bias line indicates the mean difference between the observed and linked scores; 95% of the differences are expected to fall within the upper and lower limits of agreement (LOA). The Spearman correlation coefficient indicates that there is no linear relationship between the severity of depression and the difference between the observed and linked scores at 6 months.
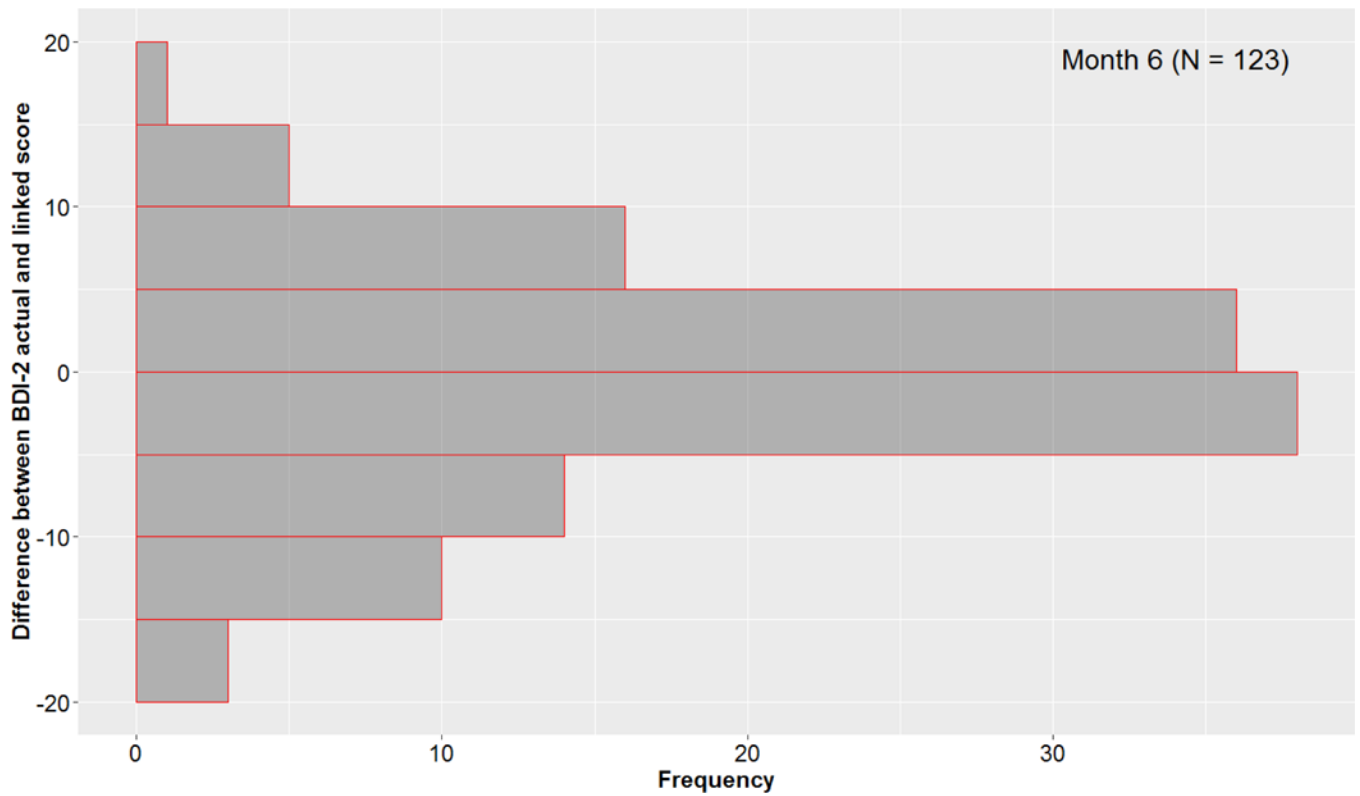
Figure 2c.
Histogram of differences between 6-month BDI-II observed and linked scores, collapsed over all levels of severity. The vertical axis represents the difference between the scores, the horizontal axis represents the number of participants, and the bars depict 5-point increments along the vertical axis. The graph shows that while extreme within-participant differences between the scores were rare, small-to-moderate differences were common. It also shows that the differences were bidirectional.

Table 1.

PROMIS Depression and BDI-II scores at baseline and 6 months.

Measure	Baseline (n = 158)			6 Months (n = 123)		
	Mean	S.D.	Range	Mean	S.D.	Range
PROMIS Depression Score	63.0	6.2	42.1 to 78.1	53.6	8.4	33.6 to 76.5
BDI-II observed score	30.2	8.5	15.0 to 54.0	13.5	9.8	0.0 to 45.0
BDI-II linked score	27.1	10.9	2.0 to 55.0	13.6	10.3	0.0 to 52.0
Difference (observed-linked)	3.1	9.3	−24.0 to 35.0	−0.2	6.9	−20.0 to 19.0