Sensitivity to change of the Beck Depression Inventory versus the Inventory of Depressive Symptoms

Meesters , Y.; Sijbrandij, J.J.; Visser, E.; Beurs, E. de

Journal of Affective Disorders 281 (2021) 338–341

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Sensitivity to change of the Beck Depression Inventory versus the Inventory of Depressive Symptoms

Ybe Meesters [a,*], Jitske J. Sijbrandij [b], Ellen Visser [a], Edwin de Beurs [c]

[a] Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[b] Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[c] Faculty of Clinical Psychology, Leiden University, Leiden, The Netherlands

ARTICLE INFO

Keywords:
assessment
depression
psychometrics
sensitivity to change
BDI-II
IDS-SR

ABSTRACT

Background: In a previous study which made a comparison between disorder-specific and generic instruments to assess outcome of treatments for depression, the Beck Depression Inventory, Second Edition (BDI-II) seemed to be more sensitive to change than the Inventory of Depressive Symptoms- Self Rating (IDS-SR).

Methods: A set with longitudinal data from Routine Outcome Monitoring (n=144) were analyzed with multilevel models with random intercepts. The sensitivity to change of two disorder-specific instruments, the BDI-II and the IDS-SR, were compared head to head.

Results: The BDI-II was more sensitive to change when measuring treatment outcome compared to the IDS-SR. The BDI-II decreases significantly more over time than the IDS-SR: the average decrease per week for the IDS-SR is -.012 (95%CI -0.015, -0.009) and for the BDI-II it is -.017 (95%CI -0.021, -0.014).

Limitations: Conclusions can only be preliminary due to a small sample size.

Conclusions: Treatment outcomes measured with questionnaires may differ depending on the degree of sensitivity to change of the instruments.

Routine outcome monitoring (ROM) has been promoted to improve the quality of mental health care. This method can be used to assess the effects of treatment intervention in an individual, between individuals, or in organizations (Kilbourne et al., 2018). It is also possible to use these results for benchmarking. Until recently, in The Netherlands, a central organisation gathered data from most mental health organizations. They made it possible to choose from a limited number of generic, mostly self-rating instruments, which could be used independently of the patient's diagnosis, in order to build a national benchmark dataset.

There is an ongoing discussion about the claim that generic instruments are the optimal choice for assessing outcome in various diagnostic groups. Some studies show that the sensitivity to change of disorder-specific instruments is superior compared to generic instruments when assessing the effectiveness of a treatment, for example in assessing the clinical course of eating disorders (Dingemans and Furth, 2017).

To investigate the claim that disorder-specific instruments are superior in assessing the effectiveness of a treatment, de Beurs et al. (2019) compared the use of generic instruments to disorder-specific instruments in the treatment of depressed patients in a multi-center study, using three datasets.

In the first dataset, concerning total scores, the Beck Depression Inventory-Second Edition (BDI-II; Beck et al., 1996, Beck et al., 2016) appeared to be more sensitive to change than the Symptom Checklist 90 (SCL-90; Derogatis, 1977). In the second dataset, the BDI-II was more sensitive to change than the Brief Symptom Inventory (BSI; Derogatis and Melisaratos, 1983). Finally, in the third dataset, the Inventory of Depressive Symptomatology-Self Rating (IDS-SR; Rush et al., 1986; Rush et al., 1996) did not show superior sensitivity to change compared to the Outcome Questionnaire 45 (OQ-45; Lambert et al., 2004). These results suggest a difference in sensitivity to change between the IDS-SR and BDI-II. As sensitivity to change is the most important characteristic of an instrument measuring treatment outcome, a comparison between the two instruments should be made.

However, in the study by de Beurs et al., the two disorder-specific instruments were administered in different samples, which precluded a direct head-to-head comparison of the IDS-SR and the BDI-II. The aim of the present study is to compare treatment results in depressed patients as assessed with the IDS-SR and the BDI-II, and to explore whether there is a difference in sensitivity to change between the two instruments.

* Corresponding author at: Department of Psychiatry, University Medical Center Groningen, PO Box 30 001, 9700 RB Groningen, The Netherlands.
E-mail address: y.meesters@umcg.nl (Y. Meesters).

https://doi.org/10.1016/j.jad.2020.12.036
Received 9 March 2020; Received in revised form 25 November 2020; Accepted 6 December 2020
Available online 9 December 2020

Methods

In the study of de Beurs et al. (2019) data of a subgroup of participants who completed the IDS-SR and the BDI-II simultaneously and repeatedly was used. This administrative dataset of the University Center for Psychiatry (UCP) of the University Medical Center Groningen (UMCG), the Netherlands, was updated with more recent patient data and used in the current study. The UCP is a specialized mental health care institution and patients were treated in different ways in outpatient as well as clinical treatments. All patients were diagnosed in an intake session by a clinician and met the criteria for a primary or secondary diagnosis of depression. Patients were selected when they completed the IDS-SR and the BDI-II simultaneously and more than once. The BDI-II and the IDS-SR were collected from January 2012 until October 2018. All data were collected by ROM several times and on a regular basis, depending on the duration of the treatment. In this study we compare the data of 2-6 consecutive measurements.

Instruments

The BDI-II (Beck et al., 1996) is a self-rating instrument, containing 21 items, with four response options (0-3) to assess the severity of depression. All items have different response options. Besides the total score (BDI-II-Tot), in the Dutch version, three subscale scores can be calculated: the cognitive subscale score (BDI-II-Cog) with 7 items, the affective subscale (BDI-II-Aff) with 5 items, and the somatic subscale (BDI-II-Som) with 9 items (Beck et al., 2016). The BDI-II has a good construct validity (Cronbachs alpha: 0.90; Brouwer et al., 2013).

The IDS-SR is a self-rating instrument containing 30 items to assess the severity of depression (Rush et al., 1986; Rush et al., 1996), also with four response options (0-3), which also vary for each item. Besides a total score based on 28 items (IDS-SR-Tot), two subscale scores can be distinguished. First, the subscale for mood and cognition (IDS-SR--Mood), 11 items, and second the subscale for anxiety/arousal (IDS-S-R-Anx), 8 items (Wardenaar et al, 2010). The IDS-SR has a good construct validity (Cronbachs alpha: 0.93; Rush et al, 1996).

Statistical analyses

The scores of both instruments were standardized on the pre-test mean and standard deviation (SD), to put all scores on a common metric. All patients with at least two valid measurements were included.

Multilevel models with random intercepts were fit to the data. Time was measured continuously in weeks since the pre-test. No random slopes were added, since the variance of the random slopes was very close to zero in exploratory analyses. First, to determine whether the BDI-II was more sensitive to change than the IDS-SR, a multilevel model with two outcomes was fitted in Mplus (Muthén and Muthén, 1998-2017). Since the BDI-II and IDS-SR both measure depression, the random intercepts (at the person level) and the residual variance (at the time level) were allowed to be correlated. A Wald test of parameter constraints was used to assess whether the change over time was significantly different for the BDI-II compared to the IDS-SR. Second, in separate multilevel models, the change over time in the subscales was assessed.

Results

In this study, 144 patients, 66 (45.8%) male, 78 (54.2 %) female were included (mean age 29.74, SD = 9.3 years).

The mean follow-up time was 56.03 weeks (SD = 31.23).

The BDI-II decreases significantly more over time than the IDS-SR. While the average decrease per week is -.012 (95%CI -0.015, -0.009) for the IDS-SR it is -.017 (95%CI -0.021, -0.014) for the BDI (Table 1). The Wald test of parameter constraints confirmed that the BDI-II decreases significantly more per week than the IDS-SR (Wald (1) = 31.2, p <.001).

Table 1
Multilevel model of time in weeks on the BDI-II and IDS-SR total score and their subscales (N = 144).

Parameter	Parameter estimate	95% Confidence interval		P-value
		Lower bound	Upper bound	
Within level				
Slope – mean IDS-SR	-0.012	-0.015	-0.009	<.001
Slope – mean BDI-II	-0.017	-0.021	-0.014	<.001
Slope IDS-SR-Mood	-0.014	-0.017	-0.01	<.001
Slope IDS-SR-Anx	-0.009	-0.011	-0.007	<.001
Slope BDI-II-Aff	-0.014	-0.017	-0.011	<.001
Slope BDI-II-Cog	-0.014	-0.018	-0.011	<.001
Slope BDI-II-Som	-0.013	-0.016	-0.01	<.001
Residual correlation IDS-SR, BDI-II	0.883	0.859	0.908	<.001
Residual variance IDS-SR	0.863	0.727	0.998	<.001
Residual variance BDI-II	1.335	1.128	1.542	<.001
Residual variance IDS-SR-Mood	1.013	0.854	1.173	<.001
Residual variance IDS-SR-Anx	0.483	0.407	0.56	<.001
Residual variance BDI-II-Aff	0.932	0.787	1.078	<.001
Residual variance BDI-II-Cog	1.082	0.914	1.251	<.001
Residual variance BDI-II-Som	0.951	0.802	1.101	<.001
Between level				
Intercept – mean IDS-SR	-0.380	-0.57	-0.189	<.001
Intercept – mean BDI-II	-0.534	-0.746	-0.321	<.001
Intercept – mean IDS-SR-Mood	-0.394	-0.589	-0.198	<.001
Intercept – mean IDS-SR-Anx	-0.235	-0.4	-0.07	0.005
Intercept – mean BDI-II-Aff	-0.416	-0.6	-0.232	<.001
Intercept – mean BDI-II-Cog	-0.471	-0.663	-0.279	<.001
Intercept – mean BDI-II-Som	-0.384	-0.562	-0.207	<.001
Correlation random intercepts IDS-SR, BDI-II	0.924	0.885	0.963	<.001
Intercept – variance IDS-SR	0.810	0.553	1.067	<.001
Intercept – variance BDI-II	0.839	0.544	1.134	<.001
Intercept – variance IDS-SR-Mood	0.786	0.523	1.049	<.001
Intercept – variance IDS-SR-Anx	0.72	0.512	0.929	<.001
Intercept – variance BDI-II-Aff	0.673	0.446	0.901	<.001
Intercept – variance BDI-II-Cog	0.701	0.458	0.944	<.001
Intercept – variance BDI-II-Som	0.578	0.367	0.79	<.001

Table note: The IDS-SR and BDI-II total scores were analysed in one multilevel model. The subscales were all analysed in separate models.

The residual within wave correlation between the IDS-SR and the BDI-II is r =.924. The person level correlation between both instruments is r = .883.

The scores on all subscales decreased significantly over time (Table 1). The lowest decline was in the IDS-SR-Anx with -.009 standard deviations per week. The IDS-SR-Mood decreased with .014 standard deviations, just as the BDI-II-Aff and BDI-II-Cog. The BDI-II-Som decreased on average with .013 standard deviations per week.

Therefore, the sensitivity to change of the BDI-II would seem to be

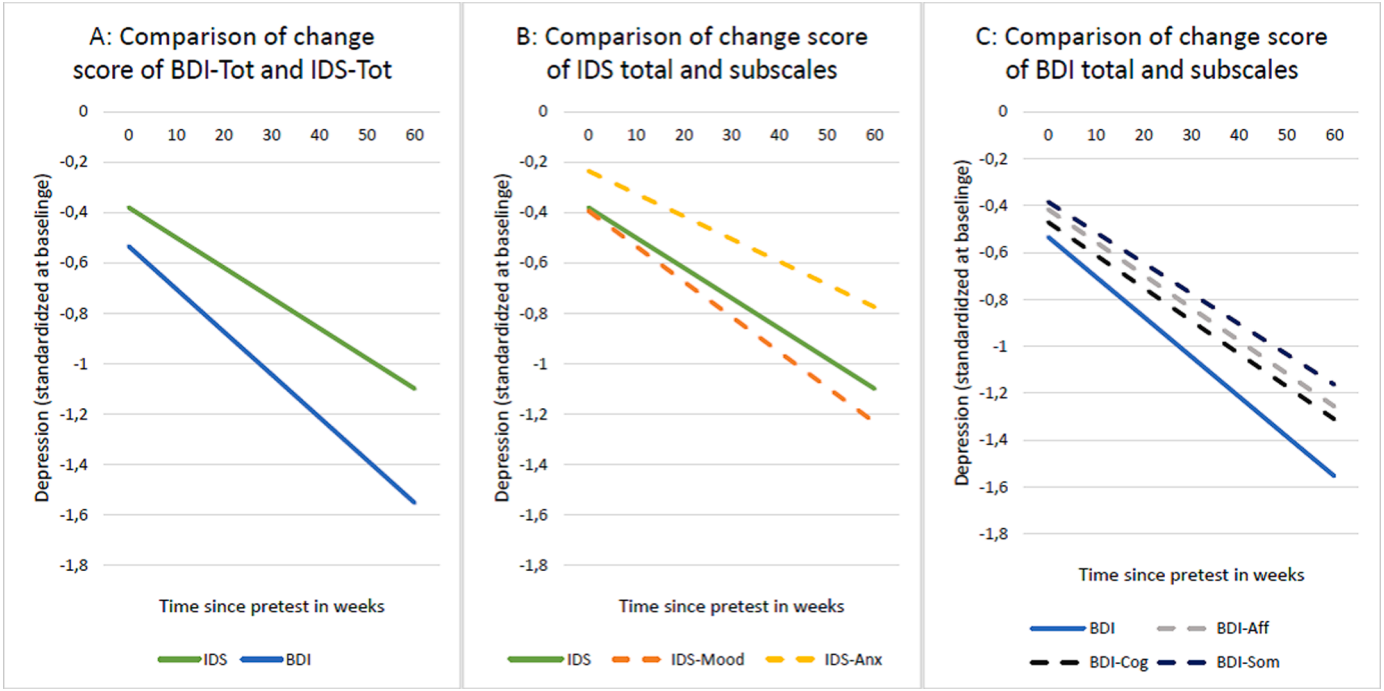

Fig. 1. Course over time of standardized scores on the IDS-SR and BDI-II and their sub-scales (n = 144). BDI-II-Tot = BDI-II total; BDI-II-Aff = BDI-II Affective; BDI-II-Cog = BDI-II Cognitive; BDI-II-Som = BDI-II Somatic; IDS-SR-Tot = IDS-SR Total; IDS-SR-Mood = IDS-SR Mood and Cognition; IDS-SR-Anx = IDS-SR Anxiety/Arousal.

superior to that of the IDS-SR. Fig. 1, shows the results of the assessments, and confirms this finding: the slope of BDI-Tot is steeper than that of IDS-SR-Tot. Moreover, the slopes for the majority of BDI subscales are steeper than the slopes of the IDS-SR subscales.

Discussion

In this study, the sensitivity to change of the BDI-II appeared to be superior to that of the IDS-SR. This finding is not in line with the existing literature. Rush et al. (1986) referred to a study about treatment effects in atypical depressions in which the IDS-SR was more sensitivity to change when assessing treatment outcome compared to other instruments. And, as shown by Kounali et al. (2016) the BDI-II rated poorly on sensitivity to change compared to the PHQ9 and other instruments. They found important differences in the sensitivity to change of several instruments.

One of the reasons to develop the IDS was the omission of some relevant questions in the other common instruments assessing the severity of depression at that time (Rush et al., 1986). Rush et al. criticized the use of the BDI (Beck et al., 1961) because of its over-valuation of cognitive items (52%), as compared to vegetative and other depressive symptoms. The BDI did not include all criteria of depression according to the DSM-IV criteria (APA, 1994; Rush et al., 1986; Rush et al. 1996). The second edition of the BDI addressed some of these criticisms (Beck et al., 1996). In our study, the most pronounced changes in scores were those on the subscales of instruments containing items about cognition. However, since the BDI-II cognitive subscale contains 7 (33.3%) items and the IDS-SR subscale mood and cognition 11 (39.3%), the difference in sensitivity to change is not explained by too much emphasis on cognitive aspects in the BDI-II.

In the study of de Beurs et al. (2019) no differences were found between the sensitivity to change of a generic instrument (OQ45) and a disorder-specific instrument (IDS-SR). Our present finding that the BDI-II had a higher sensitivity to change than the IDS-SR is in line with the previous study of de Beurs et al. (2019), indicating equal sensitivity to change of the IDS-SR compared to the OQ45, but superior sensitivity to change of the BDI-II compared to the OQ-45.

It is important to nuance our results, however. The difference was small and the within level correlation between both measures was substantial ($r = .924$). This might, in part, be due to the fact that the BDI and IDS are both self-report questionnaires and their intercorrelation may be inflated, because they are measured by the same method (i.e., self-reported). However, when two measures are intended to measure the same concept, as is the case for the BDI and IDS, intercorrelation is usually considered due to construct overlap, rather than shared method variance (Spector, 1987). Nevertheless, method variance may still account for some part of the correlation between the BDI and IDS.

As far as we are aware, this is the first longitudinal study with a head-to-head comparison of the sensitivity to change of the BDI-II and IDS-SR. Due to the rather small sample and the divergent results in the literature, our results should be considered preliminary and replication studies would be in order.

Contributors

Ybe Meesters had a role in study design, project administration, analysis and interpretation of data, and in writing original draft, reviewing and editing the final version.

Jitske J. Sijbrandij had a role analysis and interpretation of data, and in writing original draft, reviewing and editing the final version.

Ellen Visser had a role in study design, project administration, analysis and interpretation of data, and in writing original draft, reviewing and editing the final version.

Edwin de Beurs had a role in study design, analysis and interpretation of data, and in writing original draft, reviewing and editing the final version.

Declaration of Competing Interest

This research did not receive any specific grant from funding

agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

The authors are grateful to Josie Borger for the improvement of the English.

References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, 4th edn. APA, Washington, DC.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4 (6), 561–571. https://psycnet.apa.org/doi/10.1001/archpsyc.1961.01710120031004.

Beck, A.T., Steer, R.A., Brown, G.K., 1996. BDI-II Manual. Pearson, San Antonio Tx.

Beck, A.T., Steer, R.A., Brown, G.K., 2016. BDI-2-NL-R. Manual, Dutch revised edition, translation and editing: Van der Does. A.J.W. Pearson Assessment and Information BV, Amsterdam.

de Beurs, E., Visser, E., Schoevers, R., Carlier, I.V.E., van Hemert, A.M., Meesters, Y, 2019. Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three longitudinal datasets. Depress. Anxiety 36, 93–102. https://doi.org/10.1002/da.22809.

Brouwer, D., Meijer, R.R., Zevalkink, D.J., 2013. On the Factor Structure of the Beck Depression Inventory-II: G Is the Key. Psychol. Assessment 25 (1), 136–145. https://psycnet.apa.org/doi/10.1037/a0029228.

Derogatis, L.R., 1977. SCL-90: Administration, Scoring and Procedures Manual-I for the R (evised) Version. Clinical Psychometrics Research Unit. Baltimore.

Derogatis, L.R., Melisaratos, N., 1983. The Brief Symptom Inventory: an introductory report. Psychol. Med. 13, 595–605. https://psycnet.apa.org/doi/10.1017/S0033291700048017.

Dingemans, A.E., Van Furth, E.F., 2017. Het meten van verandering tijdens de behandeling van eetstoornissen: een vergelijking van algemene en specifieke vragenlijst. [Measuring changes during the treatment of eating disorders: a comparison of two types of questionnaires] Tijdschr. Psychiat 59, 278–285. http://www.tijdschriftvoorpsychiatrie.nl/assets/articles/59-2017-5-artikel-dingemans.pdf.

Kilbourne, A.M., Beck, K., Spaeth-Rublee, B., Ramanuj, P., O'Brien, R.W., Tomoyasu, N., Pincus, H.A., 2018. Measuring and improving the quality of mental health care: a global perspective. World Psychiatry 17 (1), 30–38. https://doi.org/10.1002/wps.20482.

Kounali, D.Z., Button, K.S., Lewis, G., Ades, A.E., 2016. The relative responsiveness of test instruments can be estimated using a meta-analytic approach: an illustration with treatments for depression. J. Clin. Epidemiol. 77, 68–77. https://doi.org/10.1016/j.jclinepi.2016.03.005.

Lambert, M.J., Gregersen, A.T., Burlingame, G.M., 2004. The Outcome Questionnaire-45, in: Maruish, M.E. (Ed.), The use of psychological testing for treatment planning and outcomes assessment. Instruments for adults. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, pp. 191–234.

Muthén, L.K., Muthén, B.O., 1998-2017. Mplus User's Guide, eighth ed. Muthén & Muthén, Los Angeles, CA. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf.

Rush, A.J., Giles, D.E., Schlesser, M.A., Fulton, C.L., Weissenburger, J., Burns, C., 1986. The Inventory for Depressive Symptomatology (IDS): preliminary findings. Psychiatry Res 18 (1), 65–87. https://doi.org/10.1016/0165-1781(86)90060-0.

Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol. Med. 26 (3), 477–486. https://doi.org/10.1017/s0033291700035558.

Spector, P.E., 1987. Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? Journal of Applied Psychology 72 (3), 438–443. https://doi.org/10.1037/0021-9010.72.3.438.

Wardenaar, K.J., Van Veen, T., Giltay, E.J., Den Hollander-Gijsman, M., Penninx, B.W.J.H., Zitman, F.G., 2010. The structure and dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in patients with depressive disorders and healthy controls. J. Affect. Disorders 125, 146–154. https://doi.org/10.1016/j.jad.2009.12.020.