Movement Disorders
Vol. 21, No. 5, 2006, pp. 668–672

Reliability and Validity of the Beck Depression Inventory in Patients With Parkinson's Disease

Martine Visser, PhD,[1]* Albert F.G. Leentjens, MD, PhD,[2] Johan Marinus, PhD,[1] Anne M. Stiggelbout, PhD,[3] and Jacobus J. van Hilten, MD, PhD[1]

[1]*Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands*
[2]*Department of Psychiatry, Maastricht University Hospital, Maastricht, The Netherlands*
[3]*Medical Decision-Making Unit, Leiden University Medical Center, Leiden, The Netherlands*

Abstract: We evaluated the validity, reliability, and potential responsiveness of the Beck Depression Inventory (BDI) in patients with Parkinson's disease (PD). In part 1 of the study, 92 patients with PD underwent a structured clinical interview for DSM major depression and based on this patients were considered depressed (PD-D) or nondepressed (PD-ND). Subsequently, patients filled in the BDI. In part 2, a postal survey consisting the BDI was performed in 185 PD patients and 112 controls. Test–retest reliability was assessed in 60 PD patients. The factor analysis revealed a cognitive–affective and a somatic factor. Cronbachs α for the BDI was 0.88. Mean BDI indicated significant differences ($P < 0.001$) between the PD and control group, between the PD-ND and PD-D group, and between PD-ND and control group. In part 1, the receiver operating characteristic curves showed that the area under the curve for the total BDI was 0.88. A cutoff was calculated for the BDI (14/15) that had the highest sum of sensitivity (0.71) and specificity (0.90). In part 2, the test–retest reliability for the BDI total score was 0.89 (intraclass correlation coefficient). The smallest real difference was 3.3 for the total BDI. The BDI is a valid, reliable, and potential responsive instrument to assess the severity of depression in PD. However, an adjusted cutoff is recommended. © 2006 Movement Disorder Society

Key words: Parkinson's disease; depression; assessment; reliability; validity

Depression is common in Parkinson's disease (PD; average prevalence, 40%),[1] with a large impact on the patient's quality of life.[2] Increasingly, depression is considered as one of the clinical domains of PD.[3,4] Assessment of depression in PD has frequently relied on the use of generic depression rating scales, either semistructured as the Hamilton rating scale for depression (HAM-D)[5] or self-assessed as the Beck depression inventory (BDI)[6] or the hospital anxiety and depression scale (HADS).[7] The HAM-D has a good diagnostic performance but the interview is time consuming. The HADS is a questionnaire that lacks somatic items, has satisfactory psychometric properties in PD,[8] but cannot be applied to screen or diagnose depression.[9] The BDI, like other depression scales, includes somatic items. This could decrease the psychometric properties of the BDI in PD as somatic items (of the HAM-D) were found to distinguish depressed and nondepressed PD patients worse than the nonsomatic items.[10] A lower prevalence of depressed PD patients was found in another study,[11] using only the nonsomatic items of the BDI, but no gold standard was used for comparison.

The BDI is one of the most frequently used self-rating scales for depression.[12] Despite its frequent use, the psychometric aspects of the BDI in PD have not been thoroughly evaluated. The internal consistency of the BDI is high in PD,[13] and with two adjusted cutoff scores, either the sensitivity or the specificity of the BDI is sufficient for either screening or diagnosing depression in PD.[14] The objective of this study was to evaluate the validity, reliability, and potential responsiveness of the BDI in the assessment of depression in patients with PD. The evaluation of the BDI is part of a larger research

*Correspondence to: Dr. Martine Visser, Department of Neurology, K5 Q 92, Leiden University Medical Center, P.O. Box 9600, NL-2300 RC Leiden, The Netherlands. E-mail: m.visser@lumc.nl

Received 30 May 2005; Revised 1 and 15 August 2005; Accepted 20 August 2005

Published online 31 January 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/mds.20792

project, the Scales for Outcomes in Parkinson's Disease (SCOPA), in which feasible and clinimetric sound instruments for all relevant domains in PD are selected or developed.

PATIENTS AND METHODS

Patients

PD patients fulfilled the U.K. Parkinson's Disease Society Brain Bank criteria for idiopathic PD (UKP-DSBB)[15] and were able to read and understand Dutch. Subjects without diseases of the central nervous system who were able to read and understand Dutch were eligible as controls. Partners were not allowed as controls because having a partner with PD may influence their outcomes. The local medical ethics committee approved the study protocol.

Design

This study was carried out in two parts.

Part 1.

Consecutively referred PD patients of the movement disorders clinic of the Maastricht University Hospital underwent a protocolized examination using the depression module of the structured clinical interview for DSM depression (SCID-D)[16] to verify major depressive disorder and classify patients into major-depressed (PD-D) or nondepressed (PD-ND). In addition, all patients were asked to complete the BDI. The presence of dementia was determined based on the DSM-IV criteria. This study was an extension of a previous study[14] (including 53 patients) in order to increase power for additional analyses.

Part 2.

A postal survey including the BDI was performed in 185 PD patients who visited the outpatient movement disorders clinic of the Leiden University Medical Center. Disease-specific information was obtained from patient records: Hoehn and Yahr staging (HY)[17] for disease severity, disease duration, and medication. Nonresponders were reminded by telephone after 2 weeks. Test–retest reliability was assessed in 60 PD patients, who received a second mailing 2 weeks after returning the first questionnaire. Patients provided control subjects of approximately the same age (age difference less than 10 years).

Statistical Analysis

Data were entered and analyzed using SPSS for windows 11.0. Subjects with more than 20% of the data missing were excluded from the analysis. Analyses were done with PD patients from part 1 and 2 separately or combined (total PD group).

Total PD Group.

Items were evaluated for floor and ceiling effects (≥80% response option 0 or 3) and item–total correlations (≥0.20).[18] A principal component factor analysis was performed to explore the factor structure of the BDI. Differences in items and means between the groups were analyzed using the Mann–Whitney *U* test and Student's *t* test for independent samples, respectively. The internal consistency was assessed using Cronbach's α.

PD Part 1.

The discriminative properties of the BDI were assessed by way of receiver operating characteristic (ROC) curves, which display the sensitivity and specificity to diagnose depression for different cutoffs.

PD Part 2.

Test–retest reliability for items was assessed with a weighted kappa (K_w, quadratic weights), whereas the test–retest reliability for the total BDI score was assessed using an intraclass correlation coefficient (ICC). As an indicator of potential responsiveness, the smallest real difference (SRD) was calculated using the following formula: $SRD = 1.96 \times \sqrt{2} \times SEM$,[19] where SEM is standard error of measurement: $\sqrt{\text{total variance}} \times \sqrt{(1 - ICC)}$.

RESULTS

Patients

Ninety-two PD patients participated in part 1: 21 PD patients (19%) were major-depressed and 71 were not depressed. Major-depressed and nondepressed patients did not differ significantly with respect to age, disease duration, sex, or HY stage. One hundred thirty-two PD patients (71%) responded to the mail survey. In the control group, 104 of the 112 subjects (93%) returned the BDI. Five patients and three control subjects had more than 20% missing values and their data were removed. The response rate for the test–retest assessment was 92%. The PD patients in part 1 and 2 did not differ significantly with respect to age but PD part 1 had a significant longer disease duration (Table 1). The control group was significantly younger than both PD groups and included significantly more men than the PD part 2 group. Therefore, analyses with the control group were corrected for age and sex.

TABLE 1. *Characteristics of participants in both parts of the study*

	Part 1	Part 2	
	PD_1	Controls	PD_2
n	92	101	127
Age (yr)	67.3 (10.7)	60.8 (11.7)	65.0 (10.9)
Sex (% male)	48/44 (52%)	49/52 (49%)	79/48 (62%)
Disease duration (yr)	5.8 (5.0)		9.3 (4.9)
HY stage	11/63/15/1/0		3/54/53/15/2

PD_{1+2}, Parkinson's disease patients of parts 1 and 2.
HY, Hoehn and Yahr.

Analyses in Total PD Group

Five items had floor effects but none of the items showed ceiling effects. In the control group, 14 items had floor effects. All item–total correlations exceeded 0.20 (range, 0.26–0.71). The total PD group scored significantly higher than controls on 16 of the 21 items (Table 2).

Factorial Validity.

Inspection of the scree plot indicated that a two- or three-factor structure was appropriate, explaining 40% or 45% of the variance, respectively. The content of the three-factor structure was difficult to interpret, especially the five items of the third factor were difficult to name. The two-factor structure was clearer: factor 1 included cognitive–affective items, and factor 2 included somatic items, together with the item "indecisiveness" (Table 2). Only the item "social withdrawal" clearly double-loaded. Based on this two-factor solution, a cognitive–affective (11 items) and a somatic (10 items) subscore were calculated.

Internal Consistency.

Cronbach's α for the total BDI was 0.88, for the cognitive–affective factor 0.85, and for the somatic factor 0.74.

Analyses in Part 1

Five patients in the PD-ND (7.0%) and 2 patients in the PD-D group (9.5%) fulfilled the DSM criteria for dementia. The PD-ND demented patients did not differ significantly on the BDI from the nondemented PD-ND patients ($P = 0.191$).

Criterion Validity.

Significantly higher item scores were obtained for the PD-D group compared to the PD-ND group, except for the somatic items "body image" and "sleep disturbance." The PD-ND patients scored significantly higher than control subjects on 10 items (7 somatic and 3 cognitive–affective items). Mean BDI total and subscores indicated significant differences between the total PD and control group, between the PD-ND and PD-D group, and between the PD-ND and control group (Table 3).

TABLE 2. *BDI items: median (interquartile range), test–retest reliability, and factor loadings*

Item	PD_{1+2}	Control group	κ	Factor 1*	Factor 2*
Mood[a]	0 (0,0)	0 (0,0)	0.69	0.63[b]	
Pessimism[a]	0 (0,1)	0 (0,0)	0.58	0.49[b]	0.28
Sense of failure[a]	0 (0,0)	0 (0,0)	0.38	0.83[b]	
Lack of satisfaction[a]	1 (0,1)	0 (0,1)	0.45	0.50[b]	0.30
Guilty feeling	0 (0,0)	0 (0,0)	0.31	0.87[b]	
Sense of punishment	0 (0,0)	0 (0,0)	0.72	0.74[b]	
Self-hate[a]	0 (0,0)	0 (0,0)	0.59	0.86[b]	
Self-accusations	0 (0,0)	0 (0,0)	0.44	0.58[b]	
Suicidal ideas	0 (0,0)	0 (0,0)	0.89	0.26[b]	
Crying spells[a]	0 (0,1)	0 (0,0)	0.42		0.35[b]
Irritability[a]	0 (0,1)	0 (0,1)	0.72	0.28[b]	
Social withdrawal[a]	0 (0,1)	0 (0,0)	0.41	0.40[b]	0.37
Indecisiveness[a]	1 (0,2)	0 (0,0)	0.66		0.53[b]
Body image[a]	0 (0,0)	0 (0,0)	0.64		0.49[b]
Work inhibition[a]	1 (1,2)	0 (0,1)	0.69		0.60[b]
Sleep disturbance[a]	1 (0,2)	0 (0,1)	0.86		0.35[b]
Fatigability[a]	1 (1,2)	1 (0,1)	0.78		0.53[b]
Loss of appetite[a]	0 (0,1)	0 (0,0)	0.75		0.61[b]
Weight loss[a]	0 (0,1)	0 (0,0)	0.81		0.54[b]
Somatic preoccupation[a]	1 (0,1)	0 (0,0)	0.63		0.58[b]
Loss of libido	1 (0,1)	0 (0,1)	0.81		0.59[b]

*Extraction method: principal component analysis; rotation method: oblique with Kaiser normalization; only factor loadings > 0.25 are presented.
[a]Mann–Withney U test ($P < 0.05$).
[b]The highest factor loadings.

TABLE 3. *BDI total and subscores*

	Part 1		Part 2	Parts 1 + 2
	PD-ND	PD-D	Controls	PD
n	71	21	101	219
BDI[a]	8.6 (5.0)	19.7 (9.1)	5.0 (4.0)	11.8 (8.2)
Somatic[a]	6.3 (3.4)	11.3 (4.9)	3.6 (2.7)	7.9 (4.7)
Cognitive–affective[a]	2.4 (2.2)	8.5 (6.2)	1.4 (2.1)	3.9 (4.5)

t test significant differences ($P < 0.000$) between controls and PD, controls and PD-ND, and PD-ND and PD-D.

Discriminant Validity.

ROC curves showed that the area under the curve (AUC) for the total BDI was 0.88, for the somatic 0.79, and for the cognitive–affective subscore 0.86. The optimal cutoff that had the highest sum of sensitivity (0.71) and specificity (0.90) was 14/15; this means that a score of 14 or less indicates no depression, whereas 15 or more indicates depression. Sensitivity and specificity for the optimal cutoffs of the cognitive–affective (4/5) and the somatic (6/7) subscale were 0.71/0.87 and 0.86/0.54, respectively.

Analyses in Part 2

Reliability.

The test–retest reliability for the individual items ranged from 0.31 to 0.86 (K_w; Table 2). Two items had only a fair agreement ($K_w < 0.40$). ICCs for the BDI total score and the somatic subscale were both 0.89, and for the cognitive–affective subscale 0.90.

Potential Responsiveness.

The SRD for the total BDI was 3.3; this is 5% of the maximum BDI score. The SRD for the somatic subscale was 2.8 (9% of the maximum), and for the cognitive–affective subscale 1.8 (5% of the maximum).

DISCUSSION

The BDI is a self-completed instrument for depression that was found to have good test–retest reliability and internal consistency in PD patients. Test–retest reliability of only two items was low. This variability may be related to fluctuations, but it was not assessed whether patients were *on* or *off* when completing the BDI. Five items in the PD group had floor effects. Floor effects hamper the differentiation between groups: of five items that did not differentiate between PD and controls, three items had floor effects. Although these items might not address prevalent features, they were able to differentiate between PD-D and PD-ND. The factor analysis suggested a cognitive–affective and a somatic factor, both with high internal consistency. Earlier research in university students[20] and depressive in-patients[21] reported similar factor solutions and a not very high explained variance of approximately 40%. The item "indecisiveness" loaded in our study on the somatic factor, whereas in other studies this item was cognitive–affective.[13,20] Also, the items "body image" and "crying spells" loaded on the somatic factor, but these could very well be addressed to having PD.

Compared to controls, PD-ND had significantly higher BDI scores. This suggests that PD-ND patients have mood disturbances without fulfilling the DSM-IV criteria of major depression. A limitation of this study is that patients with minor depression were allocated in the PD-ND group. On the other hand, this difference underscores the shared content of the BDI and PD, especially because controls and PD-ND differ significantly on seven somatic items and only three cognitive–affective items.

Although dementia and depression may have overlapping symptoms, the percentage of demented patients in both groups was similar. Furthermore, the demented patients in the PD-ND group did not score significantly higher than the rest of that group. In a PD population, the BDI should be used but with an adjusted cutoff score of 14/15. However, there is no ideal cutoff with both high sensitivity and specificity. The previous cutoff found by Leentjens and colleagues[14] differs slightly from the cutoff proposed in this study, which is based on a more extended population.

Major-depressed PD patients score higher than nondepressed PD patients on the BDI and on 8 of 10 somatic items, which cannot be explained by disease duration or severity. We conclude therefore that the somatic items "body image" and "sleep disturbance" assess disease characteristics and not depression. Although the HADS lacks somatic items, the internal consistency and test–retest reliability of the HADS[8] and the BDI are very similar. The HADS showed no floor or ceiling effects in PD but the BDI items are more in accordance to the DSM criteria for depression.[16] The use of the DSM criteria may also be questionable as a gold standard for depression in PD as these also include somatic items. However, for the diagnosis of depression, at least one of the symptoms must either be depressed mood or loss of interest or pleasure. Furthermore, no better criterion is currently available.

The SRD of the total BDI in PD is 3.3, implying that in follow-up studies, a difference of more than 3.3 on the BDI is indicative of a real change. A low SRD indicates high potential responsiveness because smaller changes are required to exceed measurement error. Although the

BDI has the potential to be a responsive instrument, this needs to be further explored for interventions of known efficacy.

In conclusion, the BDI is a reliable, valid, and potential responsive instrument to assess the severity of depression in PD. Although the somatic items increase the BDI score in PD, they do not decrease the internal consistency or the ability to differentiate between depressed and nondepressed PD patients. Therefore, we recommend that the BDI be used with an adjusted cutoff.

Acknowledgments: This study was supported by the Netherlands Organization for Scientific Research (project number 0940-33-021). We thank Prof. Dr. R.A.C. Roos for reviewing the manuscript.

REFERENCES

1. Cummings JL. Depression and Parkinson's disease: a review. Am J Psychiatry 1992;149:443–454.
2. Schrag A, Jahanshahi M, Quinn N. What contributes to quality of life in patients with Parkinson's disease? J Neurol Neurosurg Psychiatry 2000;69:308–312.
3. McDonald WM, Richard IH, DeLong MR. Prevalence, etiology, and treatment of depression in Parkinson's disease. Biol Psychiatry 2003;54:363–375.
4. Slaughter JR, Slaughter KA, Nichols D, Holmes SE, Martens MP. Prevalence, clinical manifestations, etiology, and treatment of depression in Parkinson's disease. J Neuropsychiatry Clin Neurosci 2001;13:187–196.
5. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62.
6. Beck AT, Ward CH, Mendelson M, Mock M, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:53–63.
7. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–370.
8. Marinus J, Leentjens AF, Visser M, Stiggelbout AM, van Hilten JJ. Evaluation of the hospital anxiety and depression scale in patients with Parkinson's disease. Clin Neuropharmacol 2002;25:318–324.
9. Leentjens AFG, Lousberg R, Verhey FRJ. The psychometric properties of the hospital anxiety and depression scale in patients with Parkinson's disease. Acta Neuropsychiatr 2001;13:83–85.
10. Leentjens AFG, Marinus J, van Hilten JJ, Lousberg R, Verhey FRJ. The contribution of somatic symptoms to the diagnosis of depressive disorder in Parkinson's disease: a discriminant analytic approach. J Neuropsychiatry Clin Neurosci 2003;15:74–77.
11. Emanuels-Zuurveen ES, Brouwer WH, Bouhuys AL. Physical symptoms of Parkinson's disease and the score on Beck's depression inventory. Tijdschr Gerontol Geriatr 1991;22:134–138.
12. Richter P, Werner J, Heerlein A, Kraus A, Sauer H. On the validity of the Beck depression inventory: a review. Psychopathology 1998;31:160–168.
13. Levin BE, Llabre MM, Weiner WJ. Parkinson's disease and depression: psychometric properties of the Beck depression inventory. J Neurol Neurosurg Psychiatry 1988;51:1401–1404.
14. Leentjens AFG, Verhey FRJ, Luijckx G-J, Troost J. The validity of the Beck depression inventory as a screening and diagnostic instrument for depression in patients with Parkinson's disease. Mov Disord 2000;15:1221–1224.
15. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry 1988;51:745–752.
16. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-IV-TR). Washington, DC: American Psychiatric Association; 2000.
17. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology 1967;17:427–442.
18. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford Medical Publications; 1995.
19. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001;10:571–578.
20. Endler NS, Rutherford A, Denisoff E. Beck depression inventory: exploring its dimensionality in a nonclinical population. J Clin Psychol 1999;55:1307–1312.
21. Schotte CK, Maes M, Cluydts R, De Doncker D, Cosyns P. Construct validity of the Beck depression inventory in a depressive population. J Affect Disord 1997;46:115–125.