Concurrent Validity of the Hamilton Depression Rating Scale and the Beck Depression Inventory versus the ICD-10 Diagnostic Criteria among Patients with Parkinson’s Disease

Marcos Serrano-Dueñas, MD, MSc; Sabrina Sevilla, MD; Paola Lastra, MD.

(1) Abnormal Movement Disorder Clinic, Neurological Service, Hospital Carlos Andrade Marín
(2) Medicine Faculty, Pontifical Catholic University of Ecuador
(3) Mathematical Department, Sciences Faculty, National Polytechnic School

Marcos Serrano-Dueñas, MD: has made a contribution either to the conception and design, to the acquisition of data, to the analysis and interpretation of the data; to drafting the article and reviewing it critically.

Sabrina Sevilla, MD and Paola Lastra, MD: has made a contribution to the acquisition of data, to drafting the article or reviewing it critically.

Marcos Serrano-Dueñas, MD; Sabrina Sevilla, MD and Paola Lastra, MD has given final approval of the version of the article to be published and can certify that no other individuals not listed as authors have made substantial contributions to the paper.

Conflict of interest: None

Abbreviations:
PD: Parkinson’s disease
UKPDSSBB: Kingdom PD Society Brain Bank criteria
ADL: Activities of Daily Living
HY: Hoehn and Yahr scale. The HY scale is universally used as a severity indicator, both in clinical practice and research, and for selection of patients to be included in clinical trials. It is made up of 6 points of increasing progression (from 0, No signs of disease, to 5, Wheelchair bound or bedridden unless aided)
SES: Schwab and England scale. SES scores are expressed in terms of percentage, in 11 steps from 100 to 0, where 100% denotes complete independence and 0% denotes bedridden with vegetative dysfunction
SPMSQ: Pfeiffer’s Short Portable Mental Status Questionnaire. The SPMSQ is a 10-item scale rated by interview, and is used to detect and estimate the severity of cognitive impairment. It explores memory (short- and long-term), orientation (time, place and self), and basic calculations. The SPMSQ is hardly influenced by educational level and, as a result, proves quite useful and reliable for screening general and elderly populations. Five errors or more are indicative of moderate or severe impairment.
UPDRS: Unified Parkinson’s Disease Rating Scale. The UPDRS, officially known as UPDRS version 3.0, is the most widely applied clinical rating scale for PD and a gold standard reference scale. UPDRS consists of the following 4 subscales: 1) Part I (4 items; scoring: 0-16 points), mentation, behaviour and mood; 2) Part II (13 items; 0-52 points), activities of daily living (ADL), which may be scored in “on” or “off” states; 3) Part III (14 items; 0-108 points), motor examination (this section produces 27 scores due to assessment of several signs in different parts of the body); and 4) Part IV (11 items; 23 points—not applied in the present study –), assessing for dyskinesias, fluctuations and other complications. Items belonging to the sections I to III score from 0 (normal) to 4 (severe), whereas scoring of the part IV is heterogeneous.

Abstract
Objective: To examine the concurrent validity of the Hamilton Depression Rating Scale and the Beck Depression Inventory for quantifying depression in patients with Parkinson’s disease, using the ICD-10 Diagnostic Criteria as the gold standard, and to determine if the somatization items considered are pertinent.
Methods: The study involved one hundred and forty consecutive PD patients –102 men and 38 women– with a mean age of 68.7 years and mean disease duration of 6.7 years. Sensitivity, specificity, positive and negative predictive values and likelihood ratios were obtained with a 95% CI. ROC Curves (AUC) were also performed.
Results: Based on ROC measurement of discriminative ability, our results suggest that both scales were poor at recognizing mild depression, somewhat better at recognizing moderate depression and adequate for distinguishing severe depression, though with poor specificity. Comparisons of HDRS-21, HDRS-12, BDI-21 and BDI-16 to determine concurrent validity all gave similar results for each depression level and no important differences between the complete scales (all 21 items) and abbreviated forms (without somatic items) were noted.
Conclusions: We conclude that both scales possess similar psychometric properties, but our results cannot be compared with those of other studies that used DSM-IV criteria as their gold standard. These observations led to the following conclusions: (1) the evaluation scales and criteria that comprise them were not designed for PD; (2) the somatic items observed in our patients were a product of PD; and (3) as the severity of the illness increased, so did the number of items that were confused as elements of depression.
**Introduction**

Depression is the most frequently observed neuropsychiatric symptom in Parkinson’s disease (PD), with a reported prevalence of 40%. In a community-based study, Tandberg et al. found major depressive disorder in 7.7% of patients, whereas among hospitalized patients major depression may be present in up to 70%.

The diagnosis of depression in the course of PD is a critical clinical goal due to the impact that this disease has on patients. Its diagnosis can be difficult, however, because of the similarity of signs and symptoms that are characteristic of both PD and depressive illness. This may be why depression is often under-diagnosed in patients with PD, as indicated in a study conducted by Shulman et al. that determined that doctors failed to diagnose the presence of depression in more than 50% of cases.

Two of the most commonly used scales to measure the severity of depression in PD are the Hamilton Depression Rating Scale (HDRS) and the Beck Depression Inventory (BDI). These scales were designed to evaluate the severity of depression in adult subjects without progressive degenerative neurological illness; they are not scales specifically designed to evaluate depression in PD. The HDRS has been questioned as a tool even in the area of depression, with critics citing two potential pitfalls in its psychometric properties: (1) excessive use of items that measure somatization, and (2) lack of unidimensionality. Even so, the HDRS was considered more useful than the BDI to measure depression in patients with PD, while other researchers concluded that the BDI could also be appropriately used in such patients.

Some studies analyzing the psychometric qualities of the HDRS and the BDI scales to quantify depression in PD patients used the DSM IV criteria for major depressive disorders as the gold standard. As far as we know, these scale characteristics have not been analyzed with respect to the ICD-10 criteria that classify depression at three levels—mild, moderate and severe—and note, as well, the presence or absence of somatic syndrome at each level. Conversely, the DSM IV criteria use only two categories: depressed and non-depressed patients.

This analytical, cross-sectional study at one point in time is designed to investigate the concurrent validity of the 21-item HDRS (HDRS-21) and the 21-item BDI (BDI-21) evaluation scales. We evaluated both scales against the ICD-10 criteria for depressive episodes, which were used as the gold standard. A second objective of this study was to determine the impact of somatization on the quality of each scale.

**Patients, Materials and Methods**

One hundred and forty (140) patients diagnosed with PD by UKPDSBB were consecutively selected from a pool of outpatients treated at the Clinic of Movement Disorders of the Carlos Andrade Marín Hospital (HCAM) Neurology Service in Quito, Ecuador. Patients who presented any of the following characteristics were excluded: severe cognitive impairment (evaluated by the Short Portable Mental Status Questionnaire of Pfeiffer with a score of over 5/10), serious concomitant illness, blindness, hypacusis or limb amputation. This study was approved by the HCAM Dept. of Research and Teaching and all patients involved provided prior written informed consent.

Information was collected through interviews and examination, and included demographic and historical data (age, gender, years of illness, treatment, L-dopa dosage). Patients were always evaluated during the ON period. All psychometric tools were applied to patients by the authors of this study.

During testing, patients were assessed using to the following scales and parameters: Unified Parkinson’s Disease Rating Scale (UPDRS: sections I, II and III); Illness Stage, as per Hoehn and Yahr Staging (H&Y); and ADL: Activities of Daily Living, as per Schwab and England (S&E).

In this study, one researcher used either the HDRS-21 (Spanish version by Bobes et al.) or the BDI-21 (Spanish version by Sanz et al.) to evaluate randomly selected patients. A week later, the alternate scale was used as an evaluation tool by a second researcher and, finally, a third researcher evaluated patients with the ICD-10 criteria. All the authors remained blind to ongoing test results during psychometric patient evaluations. The recommendations of Spitzer et al. were used during application of the depression scales and the ICD-10 depression criteria.

Internal consistency of the HDRS-21 and BDI-21 criteria (Cronbach’s α or η) was statistically analyzed, as were ROC curves used to evaluate the discriminative capacity of each scale. A cutoff point was established to determine the maximum sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+) and negative likelihood ratio (LR-), as well as their 95% confidence intervals (CI) for each level of depressive episode. We also measured the concordance level of scales using random concordance (Cohen’s Kappa).
Finally, we analyzed statistical differences between the depressed and non-depressed PD groups in the study using the Student’s t test.25-28

The somatization items removed in the HDRS-21 were early insomnia (4), middle insomnia (5), late insomnia (6), retardation (8), somatic anxiety (11), gastrointestinal symptoms (12), general somatic symptoms (13), genital symptoms (14) and weight loss (17). In the BDI-21, the following items were removed: sleep disturbance (16), fatigability (17), loss of appetite (18), weight loss (19) and loss of libido (20).

| CIE-10 CRITERIA FOR DEPRESSION (62 patients (44.28%)) |
|----------------------------------|------------------|------------------|------------------|------------------|------------------|
|                                  | Non Depressed    | Mild Depression  | Moderate Depression | Severe Depression | t Student’s (D vs ND) |
| Number                           | 78 (55.72%)      | 11 (7.85%)       | 28 (20%)          | 23 (16.43%)      |                  |
| Age (years)                      | 67.6 (10.2)      | 66 (11.8)        | 69.4 (10.09)     | 72.6 (12.4)      | 0.24              |
| Gender (F/M)                     | 24/54            | 2/9              | 5//23            | 7//16            |                  |
| Disease (in years)               | 6.2 (4.9)        | 6.8 (4.7)        | 7.7 (5.8)        | 6.9 (5.4)        | 0.08              |
| L-dopa (in years)                | 4.8 (4.6)        | 5.7 (4.8)        | 4.6 (3.1)        | 6.1 (5.5)        | 0.24              |
| L-dopa (dose mg/day)             | 706.6 (423.2)    | 738.6 (212.5)    | 733.03 (440.4)   | 711.9 (342.1)    | 0.41              |
| S&E                              | 72.8 (12.5)      | 60.9 (10.4)      | 66.07 (16.4)     | 59.1 (17.8)      | 0.0001            |
| Pfeiffer                          | 1.3 (1.6)        | 1.2 (1.3)        | 1.5 (1.7)        | 1.9 (2.2)        | 0.27              |
| HDRS                             | 15.2 (8.1)       | 19.4 (7.07)      | 22.9 (8.3)       | 31.4 (7.9)       | 0.0001            |
| BDI                              | 12.6 (6.9)       | 16 (3.2)         | 22.3 (6.4)       | 30.7 (9.4)       | 0.0001            |
| Non somatic syndrome             | 70 (100%)        | 9 (81.8%)        | 11 (39.2%)       | 1 (4.3%)         |                  |
| UPRS MS                          | 4.2 (2.3)        | 5.9 (1.7)        | 5.3 (2.2)        | 8.08 (2.06)      | 0.0001            |
| UPRS ADL                         | 17.4 (6.5)       | 19.5 (5.1)       | 15.6 (6.09)      | 24.1 (7.8)       | 0.0004            |
| UPRS MEx                         | 39.6 (11.9)      | 45.6 (8.07)      | 43.8 (11.5)      | 48.7 (13.08)     | 0.001             |
| Total                            | 61.2 (19.4)      | 71.4 (12.7)      | 68.7 (17.9)      | 80.9 (21.4)      | 0.0001            |
| H & Y                            | 1.5              | 5 (6.4%)         | -0 -             | 1 (3.6%)         | 2 (8.7%)          |                  |
|                                  | 2                | 27 (34.6%)       | 2 (18.2%)        | 5 (17.9%)        | 3 (13%)           |                  |
|                                  | 2.5              | 20 (25.6%)       | 2 (18.2%)        | 7 (25%)          | 2 (8.7%)          |                  |
|                                  | 3                | 25 (32.1%)       | 7 (63.6%)        | 13 (46.4%)       | 12 (52.2%)        |                  |
|                                  | 4                | 1 (1.3%)         | -0 -             | 2 (7.1%)         | 3 (13%)           |                  |
|                                  | 5                | -0 -             | -0 -             | -0 -             | 1 (4.3%)          |                  |

Divs ND = Depressed vs Non depressed; S&E = Schwab and Engad activities of daily living; UPRS MS = Mental Status; UPRS ADL = Activities of daily living; UPRS MEx = Motor examination
H & Y = Hoehn and Yahr Staging; = Activities of daily living; UPRS MEx = Motor examination
Mean value and standard deviation; p < 0.05 as significant

Table 1. Descriptive statistical of the patients

Results
The study sample was comprised of 140 patients, of whom 38 (27.1%) were female (Table 1). Based on H&Y stage classification, 8 patients were stage 1; 68 patients stage 2; 57 patients stage 3; and 7 patients either stage 4 or 5. According to ICD-10 criteria, 62 patients (44.2%) were found to be depressed; of those, 11 patients exhibited signs of mild depression, 28 of moderate depression and 23 of severe depression. In the non-depressive patient group, 66.6% were in H & Y stages 1 and 2. Among severely depressed
patients, 69.5% were in the more advanced stages (between 3 and 5). The somatic syndrome recognized by ICD-10 was present in 18.2% of the subjects suffering from mild depression, in 60.8% of patients with moderate depression and in 95.7% of severely depressed patients.

Depressed patients showed significantly higher UPDRS scores than non-depressed patients in evaluation sections and total score (Student's t p< 0.001). They also showed higher scores on the S&E Scale (Student's t p< 0.001) (Table 1). Both scales were found to be poor at recognizing mild depression, whether or not somatic items were considered. The HDRS-21 resulted in 0.5 (cutoff 9/10), the highest ROC recorded in the study. The best LR+ was 1.43 (cutoff 10/11), recorded by BDI-21.

The highest scores for moderate depression were recognized by BDI-21 (cutoff 13/14) and BDI-16 (cutoff 8/9). The highest LR+ of 1.93 was obtained using BDI-16, but the 95% CI range was too broad. Patients with severe depression scored a high ROC value of 0.88 on both HDRS-21 (cutoff 22/23) and BDI-21 (cutoff 17/18).

In general, both scales were found to be sensitive, with or without somatic items, but both provided poor specificity, except in cases of severe depression (HDRS-21 >0.79). Concordance values, following Kappa randomness correction, resulted in a cutoff point of 22/23 for HDRS-21, equivalent to 56% of the severely depressed patients (Table 2, Fig. 1).

Discussion
The main objective of this study was to assess the concurrent validity of the HDRS-21 and BDI-21 evaluation scales, often used to detect and measure depression in PD patients, using the ICD-10 criteria as a gold standard. Based on ROC measurement of discriminative ability, our results suggest that both scales were poor at recognizing mild depression, somewhat better at recognizing moderate depression and adequate for discerning severe depression, though with poor specificity.

A comparison of HDRS-21, HDRS-12, BDI-21 and BDI-16 determinations of Se, Sp, PPV, NPV, LR+ and LR- provided similar results for each level of depression (Table 2). No important differences were noted between the complete scales (using all 21 items) and their abbreviated forms (without somatic items). We conclude that HDRS and BDI possess similar psychometric properties; our results, however, cannot be compared with those of other studies that have used DSM-IV criteria as their gold standard.

We find that BDI-21 is inadequate for evaluating depression in PD patients. Our results are similar to those reported by Leentjens et al. in 2000,13 yet contrary to the conclusion of that same group of researchers six years later (2006), when they found BDI-21 to be a legitimate scale for depression evaluation in PD patients, stating: “The BDI is a valid, reliable, and potential responsive instrument to assess the severity of depression in PD. However, an adjusted cutoff is recommended.”29 Our results indicate that if the cut-off is raised, specificity improves, but sensitivity to distinguish between mild, moderate and severe depression in patients decreases.

We believe that the aforementioned contradictory results can be partially explained by the use of the ICD-10 criteria as the gold standard. We suggest that this scale is more appropriately used with "continuity criteria.”30 We also consider it advantageous to classify patients into groups –mild, moderate or severe– according to disease severity, since such grouping is likely to make therapeutic approaches more effective.31,32

Another conclusion drawn from our results is that the somatic items do not improve the psychometric qualities of these scales. Their use made little difference in α C values obtained and results were not influenced by the number of items in a scale.25,33 The ROC, Se, Sp, PPV, NPV, LR+ and LR- values were also only minimally modified by somatic items that did not contribute to an improvement in patient classification. It appears, therefore, that an evaluation of these somatic items is an unnecessary, timeconsuming step in a setting of high healthcare demand such as an outpatient clinic.34
Table 2. Cut point for maximun sensibility and specificity.
We thus conclude that somatic items should not be evaluated in depressed PD patients, contrary to the conclusion reached by Levin et al.,14 who proposed seven somatic items to be used in patient evaluation.

We observed that the greater the severity of the illness, the higher the possibility of suffering from somatic syndrome and the greater the disability (UPDRS) or detriment to daily activities (S&E) (p<0.001). These observations led to the following conclusions: (1) that the evaluation scales and criteria that comprise them were not designed for PD; (2) that the somatic items observed in our patients were a product of PD; and (3) as the severity of the illness increased, so did the number of items that were confused as elements of depression.

The problem of somatic items is complex. Some authors contend that this kind of symptoms may exaggerate the prevalence of depression in PD. As an example, Hoogendijk et al.35 used an exclusively diagnostic and etiological methodology that appeared to reduce the prevalence of depression in their study from 23% to 13% of their patients. In addition, the use of somatic items in depression scales for diagnosing the general population has been questioned.9,11 An international survey on depression and somatization36 concluded that the enormous variability in frequency of somatic items was determined by cultural differences rather than by the items themselves. Finally, evaluation scales that do not incorporate these items have been used convincingly to gauge therapeutic efficacy in depression among the general population.37,38

The scales in question were designed to quantify depression intensity. Following Haynes et al.,39 when an evaluation tool for an illness such as depression is used in a distinct context (situational or patient group), its validity is likely to be affected. Consequently, we are concerned about the ambiguity that arises in diagnosing patients who test as false positives, especially those receiving high scores on both scales.

We believe that some PD patients with significant levels of depression may not be properly diagnosed due to a lack of appropriate criteria and the use of an inadequate gold standard. For example, in another community-based study on the prevalence of depression in PD patients, Tandberg et al.2 found that only 7.7% of the sample met DSMIV criteria for MDD, but 24.1% had scores >18 on the BDI. Such results generate important questions: What is the patient’s actual diagnosis and should anti-depressive treatment be given? A possible explanation for these results is inappropriate content of evaluation tools that were used out of context, on patients different from those they were designed to evaluate. We think that such results may be due to PD-specific symptoms and signs erroneously attributed to depression35 and we share the belief that depression, as an illness, may need to be redefined.39

A recent publication on non-motor symptoms in PD40 showed that constipation was present in 46.7% of PD patients, nocturia in 66.7%, weight problems in 22.0%, sexual difficulties in 24.4% and insomnia in 40.6%, among other medical problems. It would seem extremely difficult to ask patients to determine for themselves if their own medical problems are symptoms of PD or of a depressive episode. Their doctors face the same diagnostic dilemma.

Finally, we believe that the development of a scale specifically designed to measure depression in PD is of vital importance. We also stress the importance of evaluating patients using diagnostic criteria when they present numerous or significant depressive symptoms.41

References

Figure 1. Area under curve (CUT point for maximum sensitivity and specificity)


Escalas Psicométricas (GVEEP). Evaluación psicométrica comparativa de las versiones en español de 6, 17 y 21 ítems de la Escala de valoración de Hamilton para la evaluación de la depresión. Med Clin (Barc) 2003; 120: 693-700.


