

# A comparison of nine scales to detect depression in Parkinson disease

## Which scale to use?

J.R. Williams, PhD  
E.S. Hirsch, MS  
K. Anderson, MD  
A.L. Bush, PhD  
S.R. Goldstein, MD  
S. Grill, MD  
S. Lehmann, MD  
J.T. Little, MD  
R.L. Margolis, MD  
J. Palanci, BS  
G. Pontone, MD  
H. Weiss, MD  
P. Rabins, MD  
L. Marsh, MD

Correspondence & reprint requests to Dr. Marsh: [laura.marsh2@va.gov](mailto:laura.marsh2@va.gov)

### ABSTRACT

**Objective:** The Methods of Optimal Depression Detection in Parkinson's Disease (MOOD-PD) study compared the psychometric properties of 9 depression scales to provide guidance on scale selection in Parkinson disease (PD).

**Methods:** Patients with PD (n = 229) from community-based neurology practices completed 6 self-report scales (Beck Depression Inventory [BDI]-II, Center for Epidemiologic Studies Depression Rating Scale-Revised [CESD-R], 30-item Geriatric Depression Scale [GDS-30], Inventory of Depressive Symptoms-Patient [IDS-SR], Patient Health Questionnaire-9 [PHQ-9], and Unified Parkinson's Disease Rating Scale [UPDRS]-Part I) and were administered 3 clinician-rated scales (17-item Hamilton Depression Rating Scale [HAM-D-17], Inventory of Depressive Symptoms-Clinician [IDS-C], and Montgomery-Åsberg Depression Rating Scale [MADRS]) and a psychiatric interview. *DSM-IV-TR* diagnoses were established by an expert panel blinded to the self-reported rating scale data. Receiver operating characteristic curves were used to estimate the area under the curve (AUC) of each scale.

**Results:** All scales performed better than chance (AUC 0.75-0.85). Sensitivity ranged from 0.66 to 0.85 and specificity ranged from 0.60 to 0.88. The UPDRS Depression item had a smaller AUC than the BDI-II, HAM-D-17, IDS-C, and MADRS. The CESD-R also had a smaller AUC than the MADRS. The remaining AUCs were statistically similar.

**Conclusions:** The GDS-30 may be the most efficient depression screening scale to use in PD because of its brevity, favorable psychometric properties, and lack of copyright protection. However, all scales studied, except for the UPDRS Depression, are valid screening tools when PD-specific cutoff scores are used. *Neurology*® 2012;78:998-1006

### GLOSSARY

**AUC** = area under the curve; **BDI** = Beck Depression Inventory; **CESD-R** = Center for Epidemiologic Studies Depression Rating Scale-Revised; **GDS** = Geriatric Depression Scale; **H&Y** = Hoehn and Yahr; **HAM-D-17** = 17-item Hamilton Depression Rating Scale; **IDS-C** = Inventory of Depressive Symptoms-Clinician; **IDS-SR** = Inventory of Depressive Symptoms-Patient; **MADRS** = Montgomery-Åsberg Depression Rating Scale; **MMSE** = Mini-Mental State Examination; **MOOD-PD** = Methods of Optimal Depression Detection in Parkinson's Disease; **NPV** = negative predictive value; **PD** = Parkinson disease; **PHQ-9** = Patient Health Questionnaire-9; **PPV** = positive predictive value; **ROC** = receiver operating characteristic; **SCID** = Structured Clinical Interview for *DSM* Disorders; **UPDRS** = Unified Parkinson's Disease Rating Scale.

Depressive syndromes affect an estimated 40% of patients with Parkinson disease (PD).<sup>1</sup> However, depressive syndromes in PD are often unrecognized or inadequately treated.<sup>2</sup> Routine use of depressive symptom rating scales may improve detection of depression in PD. In 2009, the US Preventive Services Task Force recommended use of depression scales in primary care but did not recommend a specific scale.<sup>3</sup> Similarly, no one scale is recommended in PD.<sup>4</sup>

Depression screening tools should be both sensitive and specific to the broad differential diagnosis of depressed mood in PD.<sup>5,6</sup> Although not a substitute for a diagnostic evaluation, scales

Supplemental data at [www.neurology.org](http://www.neurology.org)

Supplemental Data



From the Food and Drug Administration (J.R.W., S.R.G.), Silver Spring, MD; Departments of Psychiatry and Behavioral Sciences (J.R.W., E.S.H., S.L., J.T.L., R.L.M., J.P., G.P., P.R., L.M.) and Neurology (S.R.G., S.G., H.W., L.M.), Johns Hopkins University School of Medicine, Baltimore, MD; Department of Psychiatry (K.A.), University of Maryland Medical Center, Baltimore; Mental Health Care Line (A.L.B., L.M.), Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX; Menninger Department of Psychiatry and Behavioral Sciences and Department of Neurology (A.L.B., L.M.), Baylor College of Medicine, Houston, TX; Parkinson's and Movement Disorder Center of Maryland (S.R.G., S.G.), Baltimore, MD; Department of Psychiatry (J.T.L.), Georgetown University Medical Center, Washington, DC; and Mental Health Service (J.T.L.), Veterans Affairs Medical Center, Washington, DC.

*Study funding:* Funding information is provided at the end of the article.

*Disclosure:* Author disclosures are provided at the end of the article.

should distinguish normal emotional variability from symptoms that reflect major depression or a disabling nonmajor depressive syndrome.<sup>5,7</sup> Multiple clinician-rated depression scales (the 24-item and 17-item Hamilton Depression Rating Scale [HAM-D], the Montgomery-Åsberg Depression Rating Scale [MADRS], and the Unified Parkinson's Disease Rating Scale [UPDRS] Depression item),<sup>8–18</sup> and self-reported scales (Beck Depression Inventory [BDI]–Version I) and the 30-item and 15-item Geriatric Depression Scale [GDS]) have been shown to be valid in PD.<sup>14,15,18–24</sup> The GDS-15 and Patient Health Questionnaire-9 (PHQ-9) have also been investigated as diagnostic instruments in PD.<sup>25,26</sup>

The comparative performance of these scales is unknown. Previous studies evaluated 1 or 2 scales at a time in tertiary care samples and thus provide limited guidance. Differences in diagnostic approaches and subject characteristics across studies also prevent direct comparisons. In addition, the influence of patient characteristics on psychometric properties of depression scales in PD needs more study.

This study concurrently evaluated 9 depression scales relative to a comprehensive clinical assessment for depressive disorders in a community-based sample. Differences in performance were tested across scales and across patient subgroups. Accordingly, the present report provides additional guidance on depression scale selection in PD.

**METHODS Participants.** Subjects were recruited from 3 community-based movement disorder practices. Patients were mailed letters by their neurologist inviting participation in a study evaluating the psychometric performance of depression scales in PD (the Methods of Optimal Depression Detection in Parkinson's Disease [MOOD-PD] study). Subjects were included in the study if they met United Kingdom Brain Bank criteria for idiopathic PD, had a Mini-Mental State Examination (MMSE) score  $\geq 24$ , and spoke English fluently.<sup>27,28</sup>

A total of 269 patients met the screening criteria and completed questionnaires, which identified by self-report or informant those who 1) endorsed any degree of depression, apathy, anxiety, or irritability, 2) had been prescribed psychiatric medications, or 3) reported a history of or a current depressive disorder diagnosis. Initially, subjects endorsing criterion 1, 2, or 3 and every fourth subject not meeting criterion 1, 2, or 3 were asked to take part in diagnostic psychiatric interviews (figure 1). After only 10 of the first 143 subjects failed to qualify for diagnostic interviews using these criteria, all subsequent participants meeting the initial screening criteria were considered eligible for diagnostic interviews.

**Screening visit.** During an initial screening visit, subjects completed 6 self-reported depression scales plus other demographic and clinical measures. All scales were completed before the diagnostic visit and were presented in random order. The time to complete each self-reported scale was recorded. The MMSE provided a measure of global cognitive status.<sup>28</sup> The UPDRS Motor Subscale and Hoehn and Yahr (H&Y) staging criteria assessed motor deficits and status of disease progression, respectively, with the subject in an optimally medicated (on) state.<sup>17,29</sup>

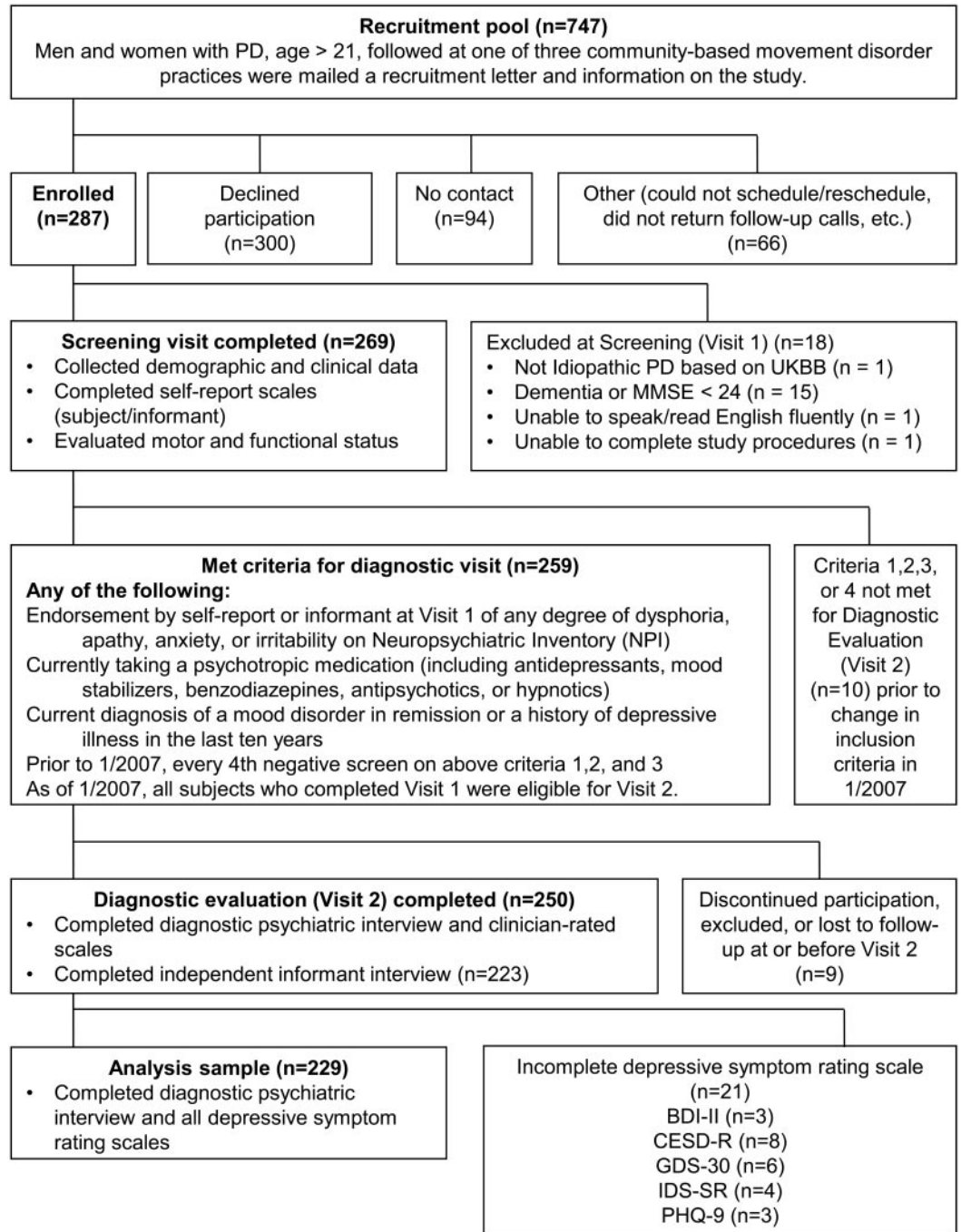
**Diagnostic visit.** A geriatric psychiatrist conducted the diagnostic examination using the Structured Clinical Interview for DSM Disorders (SCID) for DSM-IV-TR, supplemented with questions to capture psychiatric disturbances not included in the SCID (e.g., cognitive dysfunction, apathy, and impulse control disorders).<sup>30</sup> Three clinician-rated scales were rated in the context of the psychiatric interview. If an informant was available, a trained research coordinator conducted the Composite International Diagnostic Interview–Short Form and the Informant Interview for the Diagnosis of Dementia and Depression in Older Adults to obtain collateral psychiatric and cognitive information.<sup>31,32</sup> Informant information was not used to rate clinician-rated scales.

**Rating scales.** Subjects were administered 6 self-reported and 3 clinician-rated depression scales. The self-report scales (BDI-II, GDS-30, Center for Epidemiologic Studies Depression Rating Scale–Revised [CESD-R], Inventory of Depressive Symptoms–Self-Report [IDS-SR], PHQ-9, and the UPDRS Mentation, Behavior, and Mood Subscale) and the clinician-rated scales (HAM-D-17, Inventory of Depressive Symptomatology–Clinician-Rated [IDS-C], and MADRS) were selected because of previous validation in PD or because their content mapped onto DSM-IV-TR criteria for major depression.<sup>9–18,20,22,26,33–36</sup> This analysis modified the UPDRS Mentation, Behavior, and Mood Subscale by using the depression item as a self-report scale (UPDRS Depression).<sup>17</sup> An inclusive symptom attribution approach was used to rate all clinician-rated depression scales.<sup>5</sup> This approach rates all symptoms as related to depression, regardless of symptom overlap with PD or other medical conditions.<sup>5</sup>

**Psychiatric diagnoses.** Final consensus psychiatric diagnoses, based on DSM-IV-TR criteria, were established using a modification of best-estimate diagnostic procedures.<sup>33,37</sup> A panel of 6 psychiatrists with expertise in geriatric psychiatry or movement disorders reviewed each subject's history and all data collected from diagnostic interviews, informants, and medical records. The panel was completely blinded to information from self-report depression scales. An inclusive symptom attribution approach was used to diagnose depression.<sup>5</sup> Subjects were classified as actively depressed, if diagnosed with a current major depressive episode, major depressive episode in partial remission, minor depression, dysthymia, adjustment disorder with depressed mood, or depressive disorder not otherwise specified. Subjects with a depressive disorder in full remission (asymptomatic) or emotionalism (pathologic crying) were classified as not actively depressed. No instances of mood disorders due to a general medical condition or substance-induced mood disorders were observed.

**Statistical methods.** Subjects were omitted from the analysis if any depression scale items were missing. Receiver operating characteristic (ROC) curves measured the psychometric properties of the depression scales, with the consensus diagnosis serving as the gold standard.<sup>6</sup> Psychometric indices included the area under the curve (AUC), sensitivity, specificity, posi-

**Figure 1** Flow of participants through the Methods of Optimal Depression Detection in Parkinson's Disease (MOOD-PD) study



BDI = Beck Depression Inventory; CESD-R = Center for Epidemiologic Studies Depression Rating Scale-Revised; GDS-30 = 30-item Geriatric Depression Scale; IDS-SR = Inventory of Depressive Symptoms-Patient; MMSE = Mini-Mental State Examination; PD = Parkinson disease; PHQ-9 = Patient Health Questionnaire-9.

tive predictive value (PPV), and negative predictive value (NPV).<sup>6</sup> Indices were evaluated at cutoffs defined as the maximum sum of sensitivity and specificity for each scale. The maximum sum of specificity and sensitivity was chosen because these measures are not affected by prevalence rates for depression, unlike the PPV and NPV.<sup>6</sup> Internal reliability was measured by the Cronbach  $\alpha$ .<sup>6</sup>

Analyses were conducted using STATA statistical software (version 9.0). Between-group differences were evaluated using  $t$  tests

and  $\chi^2$  tests as appropriate.  $\chi^2$  tests, as implemented by the ROC-COMP command, were used to evaluate between- and within-scale AUC differences. Within-scale analyses stratified the ROC curves by gender, H&Y stage (<2.5 vs  $\geq$ 2.5), MMSE score (30–28 vs 27–26 vs 25–24), and tertiles of age, education (years), PD symptom duration (years), and UPDRS Motor score. Repeated-measures analysis of variance compared self-reported scale completion times. For post hoc pairwise AUC comparisons,  $\alpha = 0.001$  was chosen a priori. For all other analyses,  $\alpha = 0.05$  was chosen a priori.

**Standard protocol approvals, registrations, and patient consents.** The Johns Hopkins Institutional Review Board approved the study. Subjects and their informants provided written informed consent.

**RESULTS** Table 1 shows the prevalence of depression in addition to the clinical characteristics and depression scale scores of depressed and nondepressed subjects. On average, depressed subjects had fewer years of education, lower MMSE scores, and higher UPDRS

Motor scores. Depressed subjects scored higher on all depression scales. There were no statistically significant differences for any variable in table 1 between the 229 subjects included in the analyses (figure 1) and the 21 subjects who were excluded because of missing depression scale data (data not shown).

Mean scale scores by depressive subgroups are shown in figure 2. Mean scores were higher for subjects with major depression than for subjects with no active depression. Mean scores were also higher for subjects with minor depression compared with subjects with no active depression on all scales except the CESD-R, PHQ-9, and UPDRS Depression. Subjects with major depression had higher scores than subjects with minor depression on all scales except the IDS-SR.

The mean time to complete the self-report scales was as follows: BDI-II, 296.16 seconds (SD = 138.14; n = 216); CESD-R, 145.36 seconds (85.99; n = 214); GDS-30, 189.45 seconds (89.27; n = 217); IDS-SR, 594.48 seconds (297.71; n = 215); PHQ-9, 111.52 seconds (80.55; n = 214); and UPDRS-Part I, 66.98 seconds (51.46; n = 215). Completion times differed among scales in the groupwise comparison ( $F_{5,217} = 610.69, p < 0.001$ ) and in all pairwise comparisons ( $p < 0.05$ ).

The AUCs for the discrimination of actively from nonactively depressed subjects and Cronbach  $\alpha$  for each scale are provided in table 2, in addition to the sensitivity, specificity, PPV, and NPV associated with the cut-off point that maximized sensitivity and specificity. For comparison, results from other validity studies in PD are provided. The ROC curves and the sensitivity, specificity, PPV, and NPV for additional cutoff points are provided in figure e-1 on the *Neurology*<sup>®</sup> Web site at [www.neurology.org](http://www.neurology.org) and table e-1, respectively.

The AUCs were not equivalent in comparisons of all scales ( $\chi^2[8] = 25.03, p = 0.0015$ ) or in comparisons of all self-reported scales ( $\chi^2[5] = 14.09, p = 0.015$ ). The AUCs for clinician-rated scales were equivalent ( $\chi^2[2] = 1.9, p = 0.3853$ ). Pairwise post hoc comparisons are shown in table 3. At  $\alpha = 0.001$ , the UPDRS Depression had a smaller AUC than the BDI-II, HAM-D-17, IDS-C, and MADRS. The CESD-R also had a smaller AUC than the MADRS.

The AUC for each depression scale was not significantly different when stratified by age, gender, education (years), PD symptom duration (years), H&Y stage, UPDRS Motor score, or MMSE score (table e-2).

No meaningful changes in AUC values or other psychometric indices were observed in sensitivity analyses that excluded subjects with minor depression (data not shown).

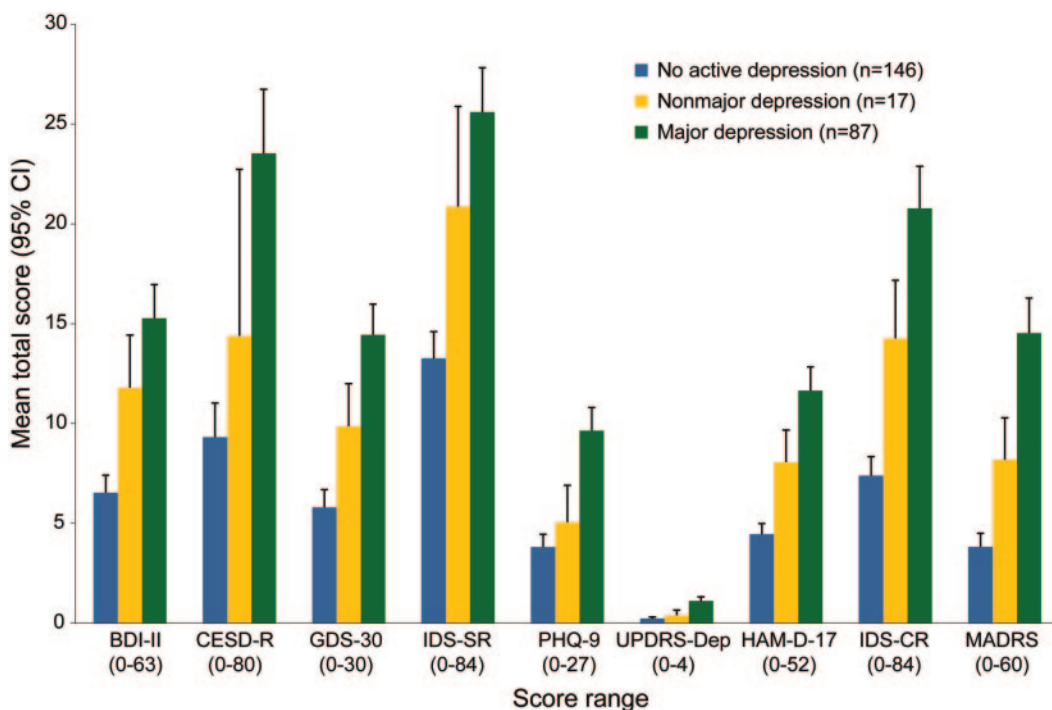
**DISCUSSION** This study compared the performance of 9 self-reported and clinician-rated depression scales in community-based patients with PD

**Table 1** Sample characteristics and scale scores

	Overall depressive disorder prevalence		
Depression, symptomatic, %	93 (40.6, 95% CI: 34.2-47.3)		
Major depression, %	78 (34.1, 95% CI: 27.9-40.6)		
Nonmajor depression, %	15 (6.6, 95% CI: 2.7-9.0)		
Depression, full remission, %	12 (5.2, 95% CI: 2.7-9.0)		
	No active depressive disorder (n = 136)	Active depressive disorder (n = 93)	p Value
Age, y, mean (SD)	66.1 (10.0)	66.0 (10.8)	<0.951
Male, n (%)	93 (68)	60 (65)	<0.542
White race, n (%)	122 (89.7)	86 (92.5)	<0.476
Education, y, mean (SD)	16.5 (3.1)	15.6 (2.6)	<0.025
PD symptom duration, y, mean (SD)	8.4 (6.7)	8.7 (6.2)	<0.721
H&Y stage, n			<0.137
I	21	10	
I½	7	0	
II	61	43	
II½	27	18	
III	13	18	
IV	5	3	
V	2	1	
Depression scale score			
UPDRS Motor, mean (SD)	15.9 (9.9) (n = 133)	21.6 (12.0) (n = 89)	<0.001
MMSE, mean (SD)	28.7 (1.3)	27.9 (1.8)	<0.001
BDI-II, mean (SD)	6.5 (5.2)	14.7 (7.4)	<0.001
CESD-R, mean (SD)	9.3 (10.1)	22.1 (15.1)	<0.001
GDS-30, mean (SD)	5.8 (5.2)	13.7 (6.8)	<0.001
IDS-SR, mean (SD)	13.3 (8.0)	24.8 (10.1)	<0.001
PHQ-9, mean (SD)	3.8 (3.8)	8.9 (5.2)	<0.001
UPDRS Depression, mean (SD)	0.2 (0.5)	1.0 (0.9)	<0.001
HAM-D-17, mean (SD)	4.5 (3.2)	11.1 (5.2)	<0.001
IDS-C, mean (SD)	7.4 (5.6)	19.7 (9.3)	<0.001
MADRS, mean (SD)	3.8 (4.0)	13.5 (7.8)	<0.001

Abbreviations: BDI-II = Beck Depression Inventory-II; CESD-R = Center for Epidemiologic Studies Depression Rating Scale-Revised; GDS-30 = 30-item Geriatric Depression Scale; H&Y = Hoehn and Yahr; HAM-D-17 = 17-item Hamilton Depression Rating Scale; IDS-C = Inventory of Depressive Symptoms-Clinician; IDS-SR = Inventory of Depressive Symptoms-Patient; PD = Parkinson disease; PHQ-9 = Patient Health Questionnaire-9; UPDRS = Unified Parkinson's Disease Rating Scale.

**Figure 2** Mean depressive symptom rating scale scores by depressive disorder subgroup



Major depression > No active depression (all scales:  $p < 0.001$ ). Nonmajor depression > No active depression (17-item Hamilton Depression Rating Scale [HAM-D-17], Inventory of Depressive Symptoms-Clinician [IDS-C]:  $p < 0.001$ ; Beck Depression Inventory-II [BDI-II], Inventory of Depressive Symptoms-Patient [IDS-SR]:  $p < 0.002$ ; Montgomery-Åsberg Depression Rating Scale [MADRS]:  $p < 0.005$ ; 30-item Geriatric Depression Scale [GDS-30]:  $p > 0.011$ ; Center for Epidemiologic Studies Depression Rating Scale-Revised [CESD-R]:  $p > 0.128$ ; Patient Health Questionnaire-9 [PHQ-9]:  $p > 0.291$ ; Unified Parkinson's Disease Rating Scale Depression [UPDRS-Dep]:  $p > 0.382$ ). Major depression > Nonmajor depression (MADRS, PHQ-9, UPDRS-Dep:  $p < 0.001$ ; HAM-D-17, IDS-C:  $p < 0.002$ ; GDS-30:  $p < 0.006$ ; CESD-R:  $p < 0.009$ ; BDI-II:  $p < 0.046$ ; IDS-SR:  $p < 0.060$ ). CI = confidence interval.

against a consensus panel depression diagnosis. All scales performed better than chance ( $AUC > 0.50$ ) in detecting depressive disorders and were not influenced by patient characteristics. The GDS-30, with its favorable psychometric properties, short administration time, and lack of copyright protection, should have broad appeal as a depression screening tool in PD; however, the BDI-II and clinician-rated scales tested also had strong psychometric properties and may be useful for depression screening.

Although all scales performed better than chance, all did not discriminate depressed patients equally well. Groupwise tests of the AUC revealed an inequality across all scales and all self-reported scales, but not across clinician-rated scales. In subsequent pairwise comparisons, all scales had similar AUCs except the CESD-R and UPDRS Depression.

Examining the scale questions provided a potential explanation for the AUC of the UPDRS Depression but not that of the CESD-R.<sup>17,35</sup> The UPDRS Depression was not developed as a single item self-report scale but rather as a clinician-rated scale.<sup>17</sup> As such, its responses are not written in lay terms. For example, a patient may not understand differences

between “sustained depression” and “sustained depression with vegetative symptoms.” In addition, a single item may not adequately assess all relevant phenomena. For example, the UPDRS Depression assumes that vegetative symptoms cannot occur without dysphoria and does not query for anhedonia. An explanation for the AUC difference between the CESD-R and MADRS is not readily apparent, because the CESD-R has questions similar to those of other scales studied.<sup>9,10,17,22,26,34–36</sup> However, the observed AUC difference is unlikely to be a type I error, because there was also a trend for differences between CESD-R and BDI-II, HAM-D-17, and IDS-C ( $p < 0.01$ ).

Most scales were sensitive to differences in symptom severity between major and nonmajor depression; however, the IDS-SR did not differentiate nonmajor from major depression and the CESD-R, PHQ-9, and UPDRS-Depression did not differentiate nonmajor depression from no active depression. Lack of sensitivity to depression severity suggests that further study is needed before the IDS-SR, CESD-R, PHQ-9, or UPDRS Depression is used to assess minor depression. These results also question whether

**Table 2** Psychometric properties of depressive symptom rating scales (n = 229): MOOD-PD and comparison screening studies

Measure	AUC	$\alpha$	Cutoff score <sup>a</sup>	Sensitivity	Specificity	PPV	NPV
<b>BDI-II</b>							
Current study	0.85	0.90	$\geq 7$	0.95	0.60	0.62	0.94
<b>CESD-R</b>							
Current study	0.79	0.92	$\geq 12$	0.72	0.70	0.62	0.79
<b>GDS-30</b>							
Current study	0.83	0.92	$\geq 10$	0.72	0.82	0.73	0.81
Reference <sup>14</sup>	0.86		$\geq 10$	0.81	0.84	0.58	0.94
Reference <sup>24</sup>	0.89		$\geq 14$	0.78	0.85	0.84	0.79
<b>IDS-SR</b>							
Current study	0.83	0.88	$\geq 14$	0.90	0.60	0.61	0.90
<b>PHQ-9</b>							
Current study	0.81	0.85	$\geq 6$	0.66	0.80	0.69	0.77
<b>UPDRS Depression</b>							
Current study	0.75	N/A	$\geq 1$	0.70	0.77	0.68	0.79
Reference <sup>16</sup>	0.79	N/A	$\geq 2$	0.66	0.81	0.81	0.66
<b>HAM-D-17</b>							
Current study	0.86	0.77	$\geq 7$	0.77	0.76	0.69	0.83
Reference <sup>11b</sup>	0.95		$\geq 14$	0.88	0.89	0.74	0.96
Reference <sup>14</sup>	N/A		$\geq 13$	0.81	0.82	0.58	0.93
<b>IDS-C</b>							
Current study	0.88	0.86	$\geq 12$	0.81	0.79	0.73	0.86
<b>MADRS</b>							
Current study	0.88	0.83	$\geq 8$	0.74	0.88	0.81	0.83
Reference <sup>11b</sup>	0.90		$\geq 15$	0.88	0.89	0.74	0.96
Reference <sup>15</sup>	0.84		$\geq 8$	0.72	0.82	0.72	0.82

Abbreviations: AUC = area under the curve; BDI-II = Beck Depression Inventory-II; CESD-R = Center for Epidemiologic Studies Depression Rating Scale-Revised; GDS-30 = 30-item Geriatric Depression Scale; HAM-D-17 = 17-item Hamilton Depression Rating Scale; IDS-C = Inventory of Depressive Symptoms-Clinician; IDS-SR = Inventory of Depressive Symptoms-Patient; MADRS = Montgomery-Åsberg Depression Rating Scale; MOOD-PD = Methods of Optimal Depression Detection in Parkinson's Disease; NPV = negative predictive value; PHQ-9 = Patient Health Questionnaire-9; PPV = positive predictive value; UPDRS = Unified Parkinson's Disease Rating Scale.

<sup>a</sup> The cutoff point that maximized the sum of sensitivity and specificity are presented for comparison with other studies, not as a recommendation for a cutoff score to be used in clinical practice.

<sup>b</sup> Results from Reijnders et al.<sup>40</sup> and Naarding et al.<sup>12</sup> were not included in the table because they reported on an expanded sample first reported by Leentjens et al.<sup>11</sup>

the aforementioned scales are appropriate to monitor treatment response; studies assessing sensitivity to change are needed to address this issue.

This study provided novel information on the time needed to complete self-report depression scales in PD. The CESD-R, GDS-30, PHQ-9, and UPDRS Depression took most subjects less than 3 minutes to complete, whereas the BDI-II took approximately 5 minutes and the IDS-SR took approximately 10 minutes. Scale length does not explain these differences because all the self-report scales except for the UPDRS Depression are composed of 20 to 30 questions.<sup>9,10,17,22,26,34-36</sup> Differing response options are a more likely explanation. The BDI-II and IDS-SR are rated in terms of symptom severity,

whereas the other scales rate the presence or frequency of symptoms.<sup>9,10,17,22,26,34-36</sup> In addition, patients may have difficulty comprehending the IDS-SR, because study staff noted that subjects often asked clarifying questions when completing this scale.

The AUC and other psychometric indices for the GDS-30, HAM-D-17, MADRS, and UPDRS Depression were similar to those previously reported in PD, but cutoff scores based on the maximum sum of sensitivity and specificity differed across studies.<sup>11,12,14-16,24</sup> In addition, cutoff scores were generally lower than suggested cutoffs scores for primary care patients with major depression.<sup>34,35,38,39</sup> This study is the first to evaluate the BDI-II, CESD-R, IDS-C, and IDS-SR in PD. One study evaluated the

**Table 3** Between-scale AUC comparisons (n = 229)<sup>a</sup>

	BDI-II	CESD-R	GDS-30	IDS-SR	PHQ-9	UPDRS-Depression	HAM-D-17	IDS-C
CESD-R	0.01							
GDS-30	0.41	0.13						
IDS-SR	0.36	0.06	0.98					
PHQ-9	0.06	0.34	0.34	0.22				
UPDRS Depression	<0.001	0.27	0.01	0.01	0.07			
HAM-D-17	0.48	0.01	0.19	0.17	0.03	<0.001		
IDS-C	0.11	0.002	0.03	0.03	0.004	<0.001	0.17	
MADRS	0.14	<0.001	0.04	0.05	0.004	<0.001	0.32	0.85

Abbreviations: AUC = area under the curve; BDI-II = Beck Depression Inventory-II; CESD-R = Center for Epidemiologic Studies Depression Rating Scale-Revised; GDS-30 = 30-item Geriatric Depression Scale; HAM-D-17 = 17-item Hamilton Depression Rating Scale; IDS-C = Inventory of Depressive Symptoms-Clinician; IDS-SR = Inventory of Depressive Symptoms-Patient; PHQ-9 = Patient Health Questionnaire-9; UPDRS-Depression = Unified Parkinson's Disease Rating Scale.

<sup>a</sup> p Values for pairwise  $\chi^2_{(1)}$  tests for AUC values are listed in table 2. A priori statistical significance was set at  $\alpha = 0.001$ .

PHQ-9 as a diagnostic instrument in PD but not for depression screening, as used in this study.<sup>25</sup>

Sampling and methodologic differences may explain the variability in reported cutoff scores. One possibility is that this community-based sample had less psychopathology compared with the previously studied tertiary care samples. Thus, lower cutoff scores may be a function of a scale score distribution whose mean is shifted to the left. Different diagnostic methods are another potential source of variability. In the absence of an objective diagnostic test for depression, the gold standard is the clinician's diagnosis; therefore, this and previous studies used psychiatrists' diagnosis of depression according to *DSM-IV-TR* criteria as the standard. However, application of *DSM-IV-TR* criteria in PD has limits that can affect diagnostic reproducibility.<sup>5</sup> In particular, the decision to attribute symptoms to PD or to a mood disorder and the thresholds used to decide when a sign or symptom is clinically relevant might vary among examiners. For this reason, the decision to exclude or include somatic items in the evaluation of patients with PD is controversial. A NIH workgroup recommended an inclusive approach, as used in this study, for symptom assessment and diagnosis to enhance sensitivity and reliability of diagnostic criteria.<sup>5</sup> This approach is supported by evidence of similar psychometric properties for the BDI-I in PD when somatic items were compared with the affective and cognitive items.<sup>23</sup> A final source of variability may be the explicit inclusion of minor depression in our case definition. However, this is not likely because exclusion of minor depression in sensitivity analyses did not change the psychometric indices appreciably.

Strengths of this study include its use of a community-based sample, a standardized diagnostic instrument (SCID), and expert consensus panel diagno-

ses. Limitations include the nonrandom sample potentially enriched for psychiatric morbidity. Patients were sequentially mailed recruitment letters inviting them to participate in a study to assess psychiatric scales. In addition, two-thirds of the subject pool elected not to participate. Although the prevalence of depression is within previously reported ranges, sampling biases are possible.<sup>1</sup> Furthermore, exclusion of patients with MMSE scores  $\leq 23$  limits generalizations to patients with significant cognitive impairment. In addition, generalization of these results may be limited by the underrepresentation of racial minorities and the high educational attainment in the sample. In addition, completion of clinician-rated scales during the psychiatric interview might have introduced a bias. To limit potential biases, each case was presented to the consensus panel by a psychiatrist other than the interviewing psychiatrist absent of any self-reported depression scale data or any clinician-rated scale scores. Furthermore, the time to administer clinician-rated scales could not be estimated because they were rated in the context of the clinical interview. Finally, self-report scale administration times do not account for the time to score or interpret them.

The ability to differentiate depressed from nondepressed patients should be the primary consideration when a scale for screening in clinical practice or research is chosen, but ease of administration is also an important consideration.<sup>6</sup> This study supports the use of select self-report scales as alternatives to clinician-rated scales. Self-report scales can be administered in waiting rooms and discussed with the patient during the clinical examination, a more practical approach for clinicians in comparison to the 15- to 20-minute interview required for clinician-rated scales.

Of the self-reported scales tested, the GDS-30 had a strong overall combination of psychometric and administration characteristics. Although the

BDI-II and GDS-30 had psychometric properties similar to those of the clinician-rated scales, the GDS-30 took less time to administer and has no copyright restrictions. Other self-reported scales either had a smaller AUC or did not distinguish between major and nonmajor depression. In particular, the UPDRS Depression performed more poorly than clinician-rated scales and the IDS-SR took considerably more time to complete.

This study does not provide explicit guidance on selection of PD cutoff points. Although some studies have used specific formulas to identify screening and diagnostic cutoff points, selection of a cutoff point should correspond to a scale's intended use and the resources available for follow-up.<sup>6,18,21</sup> In general, cutoff scores with high sensitivity are desirable to identify patients who merit further evaluation for depression; cutoff scores with high specificity are desirable when a scale is used as a substitute for a clinical interview, such as when a scale score is used as a criterion for entry into a clinical trial.

Effective screening is a critical first step toward reducing the morbidity associated with depression in PD. The GDS-30 may be the most efficient scale to use in neurology clinics because of its brevity and favorable psychometric properties. However, with the exception of the UPDRS Depression, all scales studied are valid depression screening tools when PD-specific cutoff scores are used. Furthermore, all 9 scales were unaffected by patient characteristics, suggesting that each can be administered to a wide range of patients with PD.

## AUTHOR CONTRIBUTIONS

Dr. Williams: study design, statistical analysis, interpretation of data, and drafting manuscript. E.S. Hirsch: acquisition of data, study coordination, interpretation of data, and revising manuscript. Dr. Anderson: expert panel member, interpretation of data, and revising manuscript. Dr. Bush: statistical analysis and revising manuscript. Dr. Goldstein: neurologic examination, interpretation of data, and revising manuscript. Dr. Grill: neurologic examination, interpretation of data, and revising manuscript. Dr. Lehmann: expert panel member, interpretation of data, and revising manuscript. Dr. Little: interviewing psychiatrist, expert panel member, interpretation of data, and revising manuscript. Dr. Margolis: expert panel member, interpretation of data, and revising manuscript. J. Palanci: acquisition of data, study coordination, interpretation of data, and revising manuscript. Dr. Pontone: interviewing psychiatrist, expert panel member, interpretation of data, and revising manuscript. Dr. Weiss: neurologic examination, interpretation of data, and revising manuscript. Dr. Rabins: expert panel member, interpretation of data, and revising manuscript. Dr. Marsh: study design, interviewing psychiatrist, expert panel member, study supervision, obtaining funding, interpretation of data, and revising manuscript.

## STUDY FUNDING

The MOOD-PD study was funded by NIMH grant R01-MH069666. Coauthors were also supported by NINDS grant P50-NS58377 (the Morris K. Udall Parkinson's Disease Research Center of Excellence at Johns Hopkins), NIA grant T32-AG027668, the Department of Veterans Affairs, the Donna Jeanne Gault Baumann Fund, and the Weldon Hall Trust. This article reflects the views of the authors and should not be construed to represent the Food and Drug Administration's views or policies.

## DISCLOSURE

Dr. Williams is currently employed by and owns stock in Biogen Idec. E.S. Hirsch reports no disclosures. Dr. Anderson has served on scientific advisory boards for Lundbeck, Inc. and CHDI Foundation; has received funding for travel from CHDI Foundation; serves as a Section Editor for *Current Treatment Options in Neurology*; serves as a consultant for Guidepoint Global; serves on the speakers' bureau for Lundbeck, Inc. and the FDA; and is supported by a VA Merit Award. Dr. Bush and Dr. Goldstein report no disclosures. Dr. Grill has received speaker honoraria from Medtronic, Inc. Dr. Lehmann receives research support from the NIH. Dr. Little has received research support from Neuro Hi Tech and the NIH/NIMH. Dr. Margolis has received research support from Janssen, Forest Laboratories, Inc., Amarin Corporation, Medivation, Inc., the NIH (NINDS/NIMH), and NARSAD; has received payment from AstraZeneca for expert testimony, royalties from Lippincott Williams & Wilkins for a textbook chapter and licensing fees and payments from patents related to stem cells; and payments for the development of educational presentations from Johns Hopkins University. J. Palanci reports no disclosures. Dr. Pontone has received research support from EMD Serono, ACADIA Pharmaceuticals, the NIH, Parkinson's Disease Foundation/Parkinson Study Group Mentored Clinical Research Award, and The Donna Jeanne Gault Baumann Fund. Dr. Weiss reports no disclosures. Dr. Rabins serves on a data safety monitoring board for the VA Health System; serves on the editorial boards of *International Journal of Geriatric Psychiatry and International Psychogeriatrics*; receives publishing royalties for *36-Hour Day* (Johns Hopkins Press 2011); receives research support from the NIH (NIA, NIMH); and has provided legal testimony on behalf of Janssen. Dr. Marsh serves on scientific advisory boards for the National Parkinson Foundation, American Parkinson's Disease Association, and the Parkinson Study Group; receives publishing royalties for *Psychiatric Issues in Parkinson's Disease: A Practical Guide* (Taylor & Francis, Informa, 2005); served as a consultant for Merck Serono, Boehringer Ingelheim, ACADIA Pharmaceuticals, Lundbeck, Inc., and Ovation Pharmaceuticals; received research support from Forest Laboratories, Inc., Eli Lilly and Company, Boehringer Ingelheim, the Michael J. Fox Foundation, and the NIH; and currently receives research support from the NIH, American Psychiatric Association, and the Dystonia Foundation.

Received June 14, 2011. Accepted in final form November 14, 2011.

## REFERENCES

1. 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.
2. Shulman LM, Taback RL, Rabinstein AA, Weiner WJ. Non-recognition of depression and other non-motor symptoms in Parkinson's disease. *Parkinsonism Relat Disord* 2002;8:193-197.
3. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med* 2009;151:793-803.
4. Schrag A, Barone P, Brown RG, et al. Depression rating scales in Parkinson's disease: critique and recommendations. *Mov Disord* 2007;22:1077-1092.
5. Marsh L, McDonald WM, Cummings J, Ravina B. Provisional diagnostic criteria for depression in Parkinson's disease: report of an NINDS/NIMH Work Group. *Mov Disord* 2006;21:148-158.
6. Maruish ME. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates; 2004.
7. Ravina B, Camicioli R, Como PG, et al. The impact of depressive symptoms in early Parkinson disease. *Neurology* 2007;69:342-347.



8. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. *Biol Psychiatry* 1988;23:271–284.
9. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62.
10. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382–389.
11. Leentjens AFG, Verhey FR, Lousberg R, Spitsbergen H, Wilmsink FW. The validity of the Hamilton and Montgomery-Åsberg depression rating scales as screening and diagnostic tools for depression in Parkinson's disease. *Int J Geriatr Psychiatry* 2000;15:644–649.
12. Naarding P, Leentjens AFG, van Kooten F, Verhey FRJ. Disease-specific properties of the Hamilton Rating Scale for depression in patients with stroke, Alzheimer's disease, and Parkinson's disease. *J Neuropsychiatry Clin Neurosci* 2002;14:329–334.
13. Guy W. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare/National Institute of Mental Health; 1976.
14. McDonald WM, Holtzheimer PE, Haber M, Vitek JL, McWhorter K, DeLong M. Validity of the 30-item geriatric depression scale in patients with Parkinson's disease. *Mov Disord* 2006;21:1618–1622.
15. Silberman CD, Laks J, Capitaio CF, Rodrigues CS, Moreira I, Engelhardt E. Recognizing depression in patients with Parkinson's disease: accuracy and specificity of two depression rating scale. *Arq Neuropsiquiatr* 2006;64:407–411.
16. Starkstein SE, Merello M. The Unified Parkinson's Disease Rating Scale: validation study of the mentation, behavior, and mood section. *Mov Disord* 2007;22:2156–2161.
17. Fahn S, Elton RL. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden CD, Calne DB, Goldstein M, eds. *Recent Developments in Parkinson's Disease II*. Florham Park, NJ: Macmillan Health Care Information; 1987:153–164.
18. Weintraub D, Oehlberg KA, Katz IR, Stern MB. Test characteristics of the 15-item geriatric depression scale and Hamilton depression rating scale in Parkinson disease. *Am J Geriatr Psychiatry* 2006;14:169–175.
19. Beck AT, Steer RA. *Beck Depression Inventory Manual*. San Antonio, TX: Psychological Corporation; 1987.
20. Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. In: Brink TL, ed. *Clinical Gerontology: A Guide to Assessment and Intervention*. New York: Haworth Press; 1986:165–173.
21. Leentjens AFG, Verhey FRJ, Luijckx GJ, 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.
22. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1983;17:37–49.
23. Visser M, Leentjens AF, Marinus J, Stiggelbout AM, van Hilten JJ. Reliability and validity of the Beck depression inventory in patients with Parkinson's disease. *Mov Disord* 2006;21:668–672.
24. Ertan FS, Ertan T, Kiziltan G, Uygucgil H. Reliability and validity of the Geriatric Depression Scale in depression in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2005;76:1445–1447.
25. Thompson AW, Liu H, Hays RD, et al. Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. *Parkinsonism Relat Disord* 2011;17:40–45.
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–613.
27. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992;55:181–184.
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician *J Psychiatr Res* 1975;12:189–198.
29. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967;17:427–442.
30. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute; 2002.
31. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen HU. The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). *Int J Methods Psychiatr Res* 1998;7:171–185.
32. Lewis S, Hinchcliffe K, Katona C, Livingston G. An informant interview for the diagnosis of dementia and depression in older adults (IDD-GMS). *Int J Geriatr Psychiatry* 1998;13:298–309.
33. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association; 2000.
34. Beck AT, Steer RA, Brown GK. *Manual for Beck Depression Inventory–II*. San Antonio, TX: Psychological Corporation; 1996.
35. Eaton WW, Muntaner C, Smith C, Tien A. Revision of the Center for Epidemiologic Studies Depression (CESD) scale. In: Maruish ME, ed. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:363–377.
36. Rush AJ, Giles DE, Schlessler MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): preliminary findings. *Psychiatry Res* 1986;18:65–87.
37. Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. *Arch Gen Psychiatry* 1982;39:879–883.
38. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr Ann* 2002;32:509–515.
39. Brink TL, Yesavage JA, Lum O, Heersema PH, Adey M, Rose TL. Screening tests for geriatric depression. *Clin Gerontol* 1982;1:37–43.
40. Reijnders J, Lousberg R, Leentjens AFG. Assessment of depression in Parkinson's disease: the contribution of somatic symptoms to the clinimetric performance of the Hamilton and Montgomery-Aasberg rating scales. *J Psychosom Res* 2010;68:561–565.