Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease[☆]

Alexander W. Thompson [a,*], Honghu Liu [b,c,d], Ron D. Hays [c,d,e], Wayne J. Katon [f], Rebecca Rausch [g], Natalie Diaz [h], Erin L. Jacob [i], Stefanie D. Vassar [g,j], Barbara G. Vickrey [g,j]

[a] Department of Psychiatry, Texas A&M Health Science Center College of Medicine, Temple, TX, USA
[b] UCLA School of Dentistry, Los Angeles, CA, USA
[c] UCLA Division of General Internal Medicine and Health Services Research, Department of Medicine, Los Angeles, CA, USA
[d] UCLA School of Public Health, Los Angeles, CA, USA
[e] RAND, Santa Monica, CA, USA
[f] University of Washington, Department of Psychiatry and Behavioral Sciences, Seattle, WA, USA
[g] UCLA Department of Neurology, Los Angeles, CA, USA
[h] Harbor-UCLA Department of Neurology, Los Angeles, CA, USA
[i] UCLA Department of Epidemiology, Los Angeles, CA, USA
[j] VA Greater Los Angeles Healthcare System, Parkinson's Disease Research, Education, and Clinical Center, Los Angeles, CA, USA

ARTICLE INFO

ABSTRACT

Purpose: To assess diagnostic accuracy of two self-administered depression measures compared to an interviewer-administered measure in subjects with Parkinson's disease (PD), and to analyze clinical and sociodemographic factors associated with disagreement among the three depression assessment tools.
Methods: We assessed 214 PD subjects using the Patient Health Questionnaire-9 (PHQ-9), the Geriatric Depression Scale-15 (GDS-15), and the Structured Clinical Interview for the DSM-IV depression module (SCID). Diagnostic accuracy of the PHQ-9 and GDS-15 compared to the SCID was evaluated. Multivariate logistic regression was conducted to analyze factors associated with measure disagreement. We compared item agreement between the PHQ-9 and SCID to test the hypothesis that there would be less agreement between items assessing depression symptoms overlapping with common PD symptoms, compared to items having minimal overlap with PD manifestations.
Results: Compared to SCID diagnosis of major depression, PHQ-9 sensitivity is 50% and specificity is 93%; GDS-15 sensitivity is 43% and specificity is 96%. The GDS-15 has 85% sensitivity and 79% specificity and the PHQ-9 has 54% sensitivity and 85% specificity compared to SCID diagnosis of minor or major depression. The PHQ-9 and SCID show more agreement on items unrelated to PD manifestations. Pain was the only factor associated with disagreement between the SCID and PHQ-9.
Conclusion: Compared to the PHQ-9, the GDS-15 had higher sensitivity and similar positive predictive value, suggesting it is a superior screening tool in clinical applications for PD. On future depression screening or diagnostic instruments, consideration should be given to excluding depression items overlapping with PD manifestations.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Depression may affect up to half of patients with Parkinson's disease (PD) [1], and appears to be associated with a more rapid progression of cognitive and motor impairment and a decreased health-related quality of life [2]. For that reason, easy to use screening tools for depression in PD are essential. Compared to a depression diagnosis based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) [3], the 30-item Geriatric Depression Scale [4] was found to have a high discriminant validity and internal consistency as a depression screening and diagnostic tool in patients with PD [5]. The 15-item Geriatric Depression Scale (GDS-15) [6] was found to have high discriminant validity when screening for major or minor depression in those with PD, with a sensitivity of 88% and specificity of 85% [7]. While the Patient Health Questionnaire-9 (PHQ-9) [8] is an effective depression

---

screening tool in varied medical settings [9], it has not been studied in people with PD.

The DSM-IV-TR criteria are considered the "gold standard" for depression diagnoses [3]. However, in patients with PD, "rating some of the core symptoms of depression is…difficult due to the considerable overlap of symptoms of depression and core symptoms of PD"(Schrag et al., 2007, p. 1079) [10]. Because of potential overlap, an NINDS/NIMH work group on depression in PD suggested omitting "markedly diminished interest," "psychomotor symptoms," and "diminished ability to concentrate" from depression measures [11]. Other literature suggests that sleep disturbance should not necessarily be considered a depression symptom because it is a well known PD manifestation and side effect of PD treatments [12]. Thus, measures emphasizing somatic symptoms of depression (e.g., the PHQ-9 [8]) may be vulnerable to misclassification. In contrast, an assessment measure avoiding somatic or cognitive symptoms (e.g. the GDS [13]) may more accurately assess depression in PD.

In this study, we compared the PHQ-9 and GDS-15 self-report depression measures to the "gold standard" Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID) Depression Module [14]. We hypothesized that the PHQ-9 would be more accurate and have more agreement than the GDS-15 when compared to the SCID because the PHQ-9 contains the same nine DSM-IV-based questions found in the SCID. We hypothesized that the PHQ-9 and SCID would show greater disagreement on items substantially or partially overlapping with symptoms of PD (psychomotor symptoms, concentration, sleep, and anhedonia) because while a clinical interviewer could be instructed to interpret symptoms and score the measure in a consistent way, patients' attribution of these symptoms may be more variable from one individual to another [11].

Based on data that "simpler" questionnaires may be preferred for older subjects [15] and because the GDS has been suggested as possibly valid in those with dementia [16], we hypothesized that there would be more disagreement between the GDS-15 and the other measures for subjects who were older, less educated, and had poorer performance on neuropsychological tests. Because the GDS-15 excludes somatic symptoms which may overlap with a medical disorder or general aging, there may also be greater disagreement between the GDS compared to the other depression measures among PD subjects as disease duration, motor disability and severity increases.

2. Methods

2.1. Sample and measures

The Parkinson's, Environment, and Gene (PEG) study enrolled 371 subjects between 1998 and 2006 who had been diagnosed with PD within the prior three years [17]. To follow progression of motor and non-motor manifestations of PD, 254 of the original cohort were eligible, consented, and examined from June 2007 to June 2009. The GDS-15 was collected at enrollment and follow-up, while the SCID and the PHQ-9 were added to follow-up measures. Data from the follow-up were used for these analyses. Subjects we knew did not complete all three measures on the same day were excluded, yielding a sample of 215. Of those, one subject was excluded due to mini-mental status exam (MMSE) [18] score below 15 (21% of the remaining sample (n = 45) had an MMSE score ≤ 23 with 18 subjects having a score of 23). The 214 included subjects were not different than the 40 excluded subjects on gender, age, ethnicity, or education (all p's ≥ 0.15).

At follow-up, movement disorder specialists evaluated all subjects using the Unified Parkinson's Disease Rating Scale (UPDRS) Motor score [19] and Modified Hoehn and Yahr Staging Scale [20]. Difficulty with daily functioning was assessed with the Parkinson's Disease Activities of Daily Living Scale [21]. Neuropsychological tests were collected by a research assistant (RA) trained and supported by a neuropsychologist: MMSE [18], Stroop Color Word Test [22], Boston naming test [23], Hopkins Verbal Learning Test [24], FAS verbal fluency test, and animal naming verbal fluency test [25].

The UCLA IRB (#G06-07-055) approved the study; all subjects provided informed consent.

2.2. Depression assessment measures

2.2.1. Structured Clinical Interview (SCID), Depression Module
Major and minor depression diagnoses from the SCID [14] are based on DSM-IV-TR criteria [3]. RAs, trained and supervised by a psychiatrist, conducted interviews and categorized subjects as having major, minor, or no depression. All RAs were trained to use an "inclusive" method during evaluations whereby all depressive symptoms were considered related to a depressive disorder regardless of possible overlap with PD symptoms.

2.2.2. Patient Health Questionnaire (PHQ)-9
The PHQ-9 is a self-administered depression measure that has been found to be accurate in medical settings [9]. Specifically based on DSM-IV-TR criteria, major depression is considered present if a subject reports five or more depressive symptoms with at least one symptom being anhedonia or depressed mood [8]. Minor depression is defined by the presence of two to four depressive symptoms with at least one symptom being depressed mood or anhedonia [3]. An RA could assist those patients who had difficulty marking responses.

2.2.3. Geriatric Depression Scale (GDS)-15
The GDS-15 is a 15-item yes/no questionnaire that does not focus on somatic symptoms and has no question about suicide. It was self-administered (unless a patient needed assistance marking responses). A score from five to nine was categorized as a minor depressive disorder and ten or higher represented a major depressive disorder [10].

While there was not a strict rule for the sequence of administration, typically the order was PHQ-9, GDS-15, then SCID. The PHQ-9 was mailed to the subject within two weeks of the in-person assessment so it is possible that it was completed either on or up to ten days before the in-person examination date, when the GDS-15 and SCID were administered.

2.3. Statistical analysis

Accuracy (sensitivity and specificity) of the PHQ-9 and GDS-15 compared to the SCID was determined in two ways: major depression vs. minor or no depression, and major or minor depression vs. no depression. As described above, a continuous score for the PHQ-9 was not used for determining presence of a depressive disorder.

Items from the SCID and PHQ-9 were assigned a priori into 3 groups, based on hypothesized overlap between depression and PD symptoms: A) items that substantially overlap between depression and core symptoms of PD (psychomotor changes and concentration), B) items that partially overlap between PD and depression (sleep and anhedonia), and C) items that do not or only minimally overlap with manifestations of PD (depression, guilt, appetite change, poor energy, and suicide). Disagreement was measured using the kappa statistic. For point estimate comparison, a summary statistic was created by averaging the percent disagreement across all items in each group. This allowed the PHQ-9 and SCID scores to have the same distribution so point estimates would be comparable. A weighted analysis of items took into account the number of items in each group and the degree of disagreement using the kappa statistic.

Multivariate step-wise ordinal logistic regression was completed to analyze potential factors associated with disagreement between instruments. Ordinal categories were assigned where: 0 = no depression, 1 = minor depression, and 2 = major depression. Three different dependent variables were created as measurements of disagreement: SCID minus PHQ-9; SCID minus GDS-15; and PHQ-9 minus GDS-15. Three models were created where the dependent variable was the absolute value of disagreement between measures (range 0–2). For each analysis, sociodemographic, clinical, and neuropsychological measures were included as independent variables. Adjusted odds ratios would refer to the odds that, after adjusting for all other factors, the variable in question is associated with extent of disagreement between the two measures. Bivariate associations of each independent variable with the dependent variables, and collinearity between independent variables were explored prior to multivariable modeling [26,27]. To assess the validity of using ordinal logistic regression, we tested the proportional odds assumption using the Score test. The proportional odds assumption held based on the Score test for the ordinal logistic regression models using the absolute value disagreement dependent variable. All results from regression analyses should be viewed with caution given the number of variables analyzed; a Bonferroni adjustment (0.05/19 = 0.0026) was used to interpret statistical significance of the 19 independent variables within each of the three logistic regression models.

3. Results

Participants had an average age of 72.5 years, were 58% male and 82% white or European; only 14% reported less than a high school education (Table 1). Most were married or widowed and either working or retired. About one-quarter reported moderate to very severe pain in the prior four weeks. On average, subjects had

Table 1
Clinical and sociodemographic characteristics of cohort[a] (n = 214).

	n (%) Unless specified
Age in years: mean (SD)	72.5 (9.6)
Gender (number female)	90 (42.1)
Ethnicity	
White or European American	176 (82.2)
Black or African American	1 (0.5)
Latino or Hispanic	26 (12.2)
Asian or Pacific Islander	3 (1.2)
Native American	8 (3.7)
Years of education: mean (SD)	13.8 (4.5)
Highest level of education	
Did not finish grade school (grades 1–8)	13 (6.1)
Finished grade school but no high school diploma	17 (7.9)
High school diploma	79 (36.9)
Technical or trade school diploma	25 (11.7)
College diploma	38 (17.8)
Graduate school diploma	31 (14.5)
Other	3 (1.4)
Can't remember	3 (1.4)
Refused/don't know	5 (2.3)
Marital status (n = 212)	
Never married	7 (3.3)
Married	155 (73.1)
Separated	3 (1.4)
Divorced	16 (7.6)
Widowed	31 (14.6)
Work status	
Employed	39 (18.2)
Unemployed	7 (3.3)
Retired	156 (72.9)
Other: disability (8), emeritus (1), housewife (1), temp work (1)	12 (5.6)
SF-36 v. 2 pain scale score: mean (SD)	62.2 (26.3)
Mean duration of PD diagnosis, years (SD)	5.2 (2.35)
Mean UPDRS motor score on medication (n = 189) (SD)	17.3 (10.3)
Hoehn and Yahr stage on medication (n = 185)	
Stage 0: no signs of disease	0 (0.0)
Stage 1: unilateral disease	23 (12.4)
Stage 1.5: unilateral plus axial involvement	9 (4.9)
Stage 2: bilateral disease, without impairment of balance	71 (38.4)
Stage 2.5: mild bilateral disease with recovery on pull test	47 (25.4)
Stage 3: mild to moderate bilateral disease; some postural instability; physically independent.	23 (12.4)
Stage 4: severe disability; still able to walk or stand unassisted	8 (4.3)
Stage 5: wheelchair bound or bedridden unless aided	4 (2.2)
Extent of difficulties with day-to-day activities due to Parkinson's disease, in the last 4 weeks[b]	
No difficulties	34 (15.9)
Mild difficulties	120 (56.1)
Moderate difficulties	47 (22.0)
High levels of difficulties	10 (4.7)
Extreme difficulties	3 (1.4)
Mean # medical comorbidities (SD)	1.57 (1.70)
Median MMSE score (IQR)	26 (24, 28)
On anti-depressant medication	49 (22.9)
SCID depression diagnoses	
○ Major depression	30 (14.0)
○ Minor depression	16 (7.5)
○ No depression	168 (78.5)
PHQ-9 depression diagnoses	
○ Major depression	28 (13.1)
○ Minor depression	23 (10.8)
○ No depression	163 (76.2)

Table 1 (continued)

	n (%) Unless specified
GDS-15 depression diagnoses	
○ Major depression	20 (9.4)
○ Minor depression	55 (25.7)
○ No depression	139 (64.9)
Mean (SD) GDS score	4.12 (3.5)

SF-36: Short-Form 36 Health Survey, Version 2 (0–100 possible range).
PD: Parkinson's disease.
UPDRS: Unified Parkinson's Disease Rating Scale.
MMSE: mini-mental status exam.
SCID: Structured Clinical Interview for DSM-IV-TR Axis I Disorders.
PHQ-9: Patient Health Questionnaire-9.
GDS-15: Geriatric Depression Scale-15.
[a] At time of assessment between June 2007 and June 2009.
[b] Parkinson's Disease Activities of Daily Living Scale.

been diagnosed with PD for 5.2 years at follow-up. Average UPDRS motor score was 17, and 81% were Hoehn and Yahr stage 2.5 or lower, indicating a group of subjects with relatively mild PD.

Based on SCID depression assessment, 14% met criteria for major depression and 7.5% for minor depression (Table 1). Based on the PHQ-9, 13% reported major depression and 11% minor depression. A larger number of subjects met criteria for minor but not major depression according to GDS-15 cutoffs (9% major depression and 26% minor depression; Table 1).

The PHQ-9 has a sensitivity of 54% and a specificity of 85% for any depressive disorder relative to the SCID (Table 2). The GDS-15 is more sensitive for any depressive disorder (85%) but slightly less sensitive than the PHQ-9 for differentiating major depression (43% vs. 50%).

Both tools agree with the SCID depression assessment more than with each other (SCID/PHQ-9 weighted kappa 0.4 (95% CI: 0.26–0.54); SCID/GDS-15 weighted kappa 0.5 (95% CI: 0.39–0.6); PHQ-9/GDS-15 weighted kappa 0.37 (0.25–0.49)), but neither has more than fair agreement with the SCID assessment (Table 3). There were 21 subjects with either minor or major depression based on structured interview found to have no depression with the PHQ-9. However, the PHQ-9 reported a depressive disorder in 26 (12%) subjects found to have no depressive disorder on clinical interview. The GDS-15 appears less likely to miss a diagnosis of any SCID depressive disorder, though it described 36 (17%) subjects as having a depressive disorder (31 minor, 5 major) when none was found on clinical interview.

Table 4 details the results of the item-by-item comparison between the SCID depression module and the PHQ-9. Three items having the highest kappas (=0.46–0.53) (depression, low energy, and guilt) were in the group of items (Group "C") we hypothesized a priori as not or only minimally overlapping with PD manifestations. Similarly, in terms of simple percentage agreement between the two tools, all three (depression, guilty, thoughts of suicide) items having highest percent agreement were in the group of items (Group "C") that we hypothesized a priori as not overlapping with PD manifestations. Paired t-test analysis showed a significant difference (p = 0.02) between the groups A and B items (items that substantially or partially overlap with symptoms of PD) and the group C items, with no difference between items in groups A and B. The primary finding is that there appears to be a modest difference between the PHQ-9 and SCID around depression items that experts judge to substantially or partially overlap with symptoms of PD and those that do not or only minimally overlap with manifestations of PD.

After Bonferroni adjustment, there is an association between SF-36 Pain Scale score and disagreement between the SCID and PHQ-9, but no other associations were observed (Table 5).

Table 2
Psychometric properties of the PHQ-9 and GDS-15 as depression assessment tools (n = 214).

Accuracy relative to SCID depression module	PHQ-9		GDS-15	
	Major depression (vs. minor or no depression)	Minor or major depression (vs. no depression)	Major depression (vs. minor or no depression)	Minor or major depression (vs. no depression)
Sensitivity (95% CI)	(15/30) 50.0 (31.3–68.7)	(25/46) 54.3 (39.0–69.1)	(13/30) 43.3 (25.5–62.6)	(39/46) 84.8 (71.1–93.7)
Specificity (95% CI)	(171/184) 92.9 (88.2–96.2)	(142/168) 84.5 (78.2–89.6)	(177/184) 96.2 (92.3–98.5)	(132/168) 78.6 (71.6–84.5)
Positive predictive value (95% CI)	(15/28) 53.6 (33.9–72.5)	(25/51) 49.0 (34.8–63.4)	(13/20) 65.0 (40.8–84.6)	(39/75) 52.0 (40.1–63.7)
Negative predictive value (95% CI)	(171/186) 91.9 (87.1–95.4)	(142/163) 87.1 (80.9–91.8)	(177/194) 91.2 (86.3–94.8)	(132/139) 95.0 (89.9–98.0)
Percent correctly classified (95% CI)	(186/214) 86.9 (81.6–91.1)	(163/214) 78.0 (69.9–81.7)	(190/214) 88.8 (83.8–92.7)	(171/214) 79.9 (73.9–85.1)

4. Discussion

4.1. Accuracy of the PHQ-9 and GDS-15 compared to the SCID

The PHQ-9 and GDS-15 are both considered reliable and valid depression measures for many patient populations. Both the long and short forms of the GDS perform well as depression assessment tools in PD [5,7]. As a diagnostic tool for depression, we expected the PHQ-9 would be more accurate than the GDS-15 when compared to the SCID because the PHQ-9 and SCID have the same questions. The PHQ-9 and GDS-15 had 53.6% and 65% positive predictive value respectively in identifying major depression, indicating that between one half and one third of patients will be false positives on these questionnaires. Neither tool appears very sensitive at identifying a major depressive disorder, further supporting the primary clinical utility of these instruments as screening tools. In that role, as a screening tool for any depressive disorder, the GDS appears better with its higher sensitivity and

Table 3
Diagnostic agreement among SCID, PHQ-9, and GDS-15 (n = 214).

		SCID		
		No depression	Minor depression	Major depression
PHQ-9	No depression	**142**	10	11
	Minor depression	15	**4**	4
	Major depression	11	2	**15**

Weighted kappa (95% CI) = 0.40 (0.26, 0.54)
Simple kappa (95% CI) = 0.34 (0.21, 0.47)

		SCID		
		No depression	Minor depression	Major depression
GDS-15	No depression	**132**	4	3
	Minor depression	31	**10**	14
	Major depression	5	2	**13**

Weighted kappa (95% CI) = 0.50 (0.39, 0.60)
Simple kappa (95% CI) = 0.40 (0.29, 0.50)

		PHQ-9		
		No depression	Minor depression	Major depression
GDS-15	No depression	**122**	11	6
	Minor depression	34	**10**	11
	Major depression	7	2	**11**

Weighted kappa (95% CI) = 0.37 (0.25, 0.49)
Simple kappa (95% CI) = 0.29 (0.18, 0.40)

negative predictive value, particularly for application in a clinical setting. Having a simple screening tool to help identify those with clinically important depressive symptoms (minor or major depression) is important, and the GDS-15 appears superior to the PHQ-9 in that capacity. Reasons for this may be the limited number of somatic questions on the GDS-15 and an easier response scale.

4.2. Agreement between PHQ-9 and SCID items

Patients and clinical interviewers were more likely to answer questions similarly if the question related to a depressive symptom did not overlap with PD manifestations. However, the data do not demonstrate overall strong agreement by self-report vs. trained interviewer assessment. This suggests there is something quite different in a self-report depression assessment and a clinical assessment, even when the same questions are being asked.

We found evidence to support our hypothesis that the PHQ-9 and SCID would have more disagreement on items overlapping with symptoms of PD. Although we made a distinction between depression items substantially overlapping with symptoms of PD and those partially overlapping, the analysis suggests that there is no clear difference between these two groups of items. In contrast, we found significant differences relative to items with no or minimal PD overlap.

One possible reason for disagreement between the self-report PHQ-9 and the SCID on these items is that because of the number of patients, there may have been more variability around the interpretation of the self-report items, whereas there were only a few interviewers and they were trained to interpret clinical information in a standardized way. Overall, these results provide preliminary support for the recommendation of the NINDS/NIMH work group recommendation that items such as anhedonia, psychomotor changes, and trouble concentrating be excluded from screening tools and diagnostic assessments of depression in PD [11].

4.3. Factors affecting disagreement between assessment tools

Because there is a fundamental difference between self-report measures and clinical interviews to assess depression, disagreement seems inevitable. When researching depression in PD, it is important to understand the factors that might affect that difference because self-report measures are needed for large-scale epidemiologic research efforts. Factors such as age, pain, and other medical comorbidities may be associated with variation in accuracy or agreement between depression assessment measures used in PD research.

A study comparing the Kessler-10 measure for psychological distress and the Composite International Diagnostic Interview

Table 4
Item-by-item comparison of the SCID depression module & PHQ-9 (n = 214).

	Kappa	% Agreement	Weighted group summary statistic
Group A & B			0.23 (0.22)
Group A: depression items that substantially overlap with symptoms of PD			0.23 (0.32)
Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual.	0.34	77.0	
Trouble concentrating on such things as reading the newspaper or watching televisions	0.41	77.6	
Group B: depression items that partially overlap with symptoms of PD			0.23 (0.31)
Trouble falling or staying asleep or sleeping too much (sleep disturbance)	0.45	73.4	
Little interest or pleasure in doing things (anhedonia)	0.35	81.3	
Group C: depression items that do not or only minimally overlap with symptoms of PD			0.17 (0.20)
Feeling down, depressed or hopeless (depression)	0.46	84.6	
Feeling tired or having little energy (low energy)	0.49	74.7	
Poor appetite or overeating	0.32	74.2	
Feeling bad about yourself — or that you are a failure or have let yourself or your family down (guilt)	0.53	86.9	
Thoughts that you would be better off dead or of hurting yourself in some way (thoughts of suicide)	0.41	92.1	

	Group A minus B	A minus C	B minus C	A&B minus C
Paired t-test — weighted score	Diff[a] = 0.000 p = 0.99	Diff = 0.052 p = 0.02	Diff = 0.052 p = 0.02	Diff = 0.052 p = 0.02

[a] Diff: difference in percentage disagreement between groups.

(CIDI) found that subjects over age 65 years had more difficulty with complex CIDI questions, leading to underreporting of depression. The authors concluded that "simpler" measures should be used for older patients [15]. This argument could be made for patients with lower education levels and cognitive difficulties. There is no guidance, however, around what defines a "simple" measure. When comparing the GDS-15 to the PHQ-9, it may be possible to conclude that the GDS-15 is less complex. The GDS-15 contains relatively short questions requiring yes/no answers. In contrast, the PHQ-9 asks a respondent to give an answer of "not at all" "several days" "more than half the days" or "nearly every day" to more lengthy questions. An analysis in which ROC curves for the GDS-15 were generated for three different age categories of PD patients found no difference in its performance across those age groups, a finding consistent with this idea if it is assumed that the GDS-15 is a "simpler" measure that performs well in older adults [28]. Contrary to our hypotheses, we did not find more disagreement between the GDS-15 and PHQ-9/SCID when subjects were older, had less education, more PD motor symptoms, and more cognitive problems.

Also, the GDS-15 may perform differently in those with chronic pain or medical problems. In one study, 10 items on the GDS-30

Table 5
Multivariate odds ratios for factors associated with disagreement between depression diagnosis categories on the SCID, PHQ-9, and GDS-15.[a]

	SCID minus PHQ-9 disagreement (n = 153)[b]		SCID minus GDS-15 disagreement (n = 154)		PHQ-9 minus GDS-15 disagreement (n = 156)	
	Odds-ratio	p-Value	Odds-ratio	p-Value	Odds-ratio	p-Value
Age	0.96	0.15	0.98	0.48	0.96	0.06
Gender — male	1.34	0.59	1.64	0.35	1.65	0.30
Ethnicity — White	1.04	0.96	0.49	0.19	0.50	0.19
Education	1.03	0.96	1.05	0.93	0.81	0.65
Marital status — married	1.06	0.91	0.68	0.46	1.27	0.65
SF-36 Pain Scale score	0.97	0.002	0.99	0.13	0.98	0.01
Duration of PD (years)	0.98	0.83	1.09	0.38	0.94	0.46
UPDRS motor score	1.00	0.97	1.05	0.04	1.05	0.03
# Medical comorbidities	1.11	0.50	1.11	0.48	1.03	0.84
On anti-depressant medication	0.71	0.57	1.75	0.29	1.82	0.24
MMSE	0.95	0.59	1.02	0.82	1.01	0.86
Stroop color total time	0.97	0.16	1.00	0.98	0.99	0.38
Stroop word total time	1.01	0.55	0.97	0.18	1.00	0.99
Stroop interference total time	1.08	0.003	1.07	0.008	1.01	0.65
Verbal fluency test — FAS	0.98	0.50	0.97	0.18	1.00	0.89
Verbal fluency test — animal	1.09	0.12	1.01	0.88	1.03	0.61
Boston naming test — total score	0.98	0.81	1.03	0.72	0.83	0.02
HVLT (verbal learning) — total recall score	0.95	0.31	0.99	0.88	0.99	0.79
# Days between assessments	1.15	0.43	0.98	0.53	0.88	0.16

MMSE: mini-mental status exam.
HVLT: Hopkins Verbal Learning Test.
[a] Dependent variable is non-directional with range 0–2; odds ratios refer to the odds that, after adjusting for all other factors listed, the variable in question is associated with extent of disagreement between the two measures.
[b] Sample size is lower than n = 214 because of missing values for independent variables.

were answered differentially by those in pain [29]. Five of those items are included on the GDS-15. Therefore, the GDS-15, despite fewer somatic items compared to the PHQ-9, may also lead to misdiagnoses of depression in the setting of somatic problems.

4.4. Limitations

One limitation is that we used trained RAs rather than expert psychiatric evaluation to diagnose major depression using the SCID. Research assistants were not blinded to PHQ-9 and GDS-15, which may have led to misclassification on the SCID. The PHQ-9 may have been completed before the SCID and GDS-15, though it was in a narrow window of time (maximum 10 days) prior to the in-person visit, minimizing possible bias. The sample used for this study consisted of those with relatively mild PD, limiting generalization of findings to patient populations more clinically diverse than this sample. In addition, subjects with MMSE scores below 24 were included in the study; self-report measures in those with cognitive impairment may be invalid leading to bias in our accuracy measurements.

4.5. Conclusions

Compared to the SCID, the PHQ-9 and GDS-15 have similar accuracy at assessing major depression. However, the GDS-15 appears superior to the PHQ-9 as a clinical screening tool for any depressive disorder in PD. Future research should examine the most effective cut-points for identifying any depressive disorder in PD using tools that include and exclude items with significant PD symptom overlap.

Acknowledgements

References

[1] Lemke MR. Depressive symptoms in Parkinson's disease. Eur J Neurol 2008 Apr;15(Suppl. 1):21–5.
[2] Burn DJ. Beyond the iron mask: towards better recognition and treatment of depression associated with Parkinson's disease. Mov Disord 2002 May;17(3):445–54.
[3] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Text revision. 4th ed. Washington, DC: American Psychiatric Association; 2000.
[4] Yesavage JA. Geriatric depression scale. Psychopharmacol Bull 1988;24(4):709–11.
[5] 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 Oct;76(10):1445–7.
[6] Sheikh J, Yesavage J. Geriatric depression scale (GDS): recent evidence and development of a shorter version. In: Brink T, editor. Clinical gerontology: a guide to assessment and intervention. New York: The Haworth Press; 1986. p. 165–73.
[7] 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 Feb;14(2):169–75.
[8] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606–13.
[9] Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the patient health questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596–602.
[10] Schrag A, Barone P, Brown RG, Leentjens AF, McDonald WM, Starkstein S, et al. Depression rating scales in Parkinson's disease: critique and recommendations. Mov Disord 2007 Jun 15;22(8):1077–92.
[11] 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 Feb;21(2):148–58.
[12] Chaudhuri KR, Schapira AH. Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment. Lancet Neurol 2009 May;8(5):464–74.
[13] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37–49.
[14] First M, Spitzer R, Gibbon M, Williams J. Structured clinical interview for DSM-IV-TR axis I disorders − patient edition (SCID-I/P, 4/2005 revision). New York: Biometrics Research Department, New York State Psychiatric Institute; 2005.
[15] O'Connor DW, Parslow RA. Different responses to K-10 and CIDI suggest that complex structured psychiatric interviews underestimate rates of mental disorder in old people. Psychol Med; 2008 Dec 2:1–5.
[16] Feher EP, Larrabee GJ, Crook 3rd TH. Factors attenuating the validity of the geriatric depression scale in a dementia population. J Am Geriatr Soc 1992 Sep;40(9):906–9.
[17] Jacob EL, Gatto NM, Thompson A, Bordelon Y, Ritz B. Occurrence of depression and anxiety prior to Parkinson's disease. Parkinsonism Relat Disord 2010;16(9):576–81. Epub 2010 Aug 1.
[18] 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 Nov;12(3):189–98.
[19] Fahn S, Elton R, Members of the UPDRS Development Committee. The unified Parkinson's disease rating scale. In: Fahn S, Marsden C, Calne D, Goldstein M, editors. Recent developments in Parkinson's disease. Florham Park, NJ: Macmillan Healthcare Information; 1987. p. 153–63.
[20] Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology 1967 May;17(5):427–42.
[21] Hobson JP, Edwards NI, Meara RJ. The Parkinson's disease activities of daily living scale: a new simple and brief subjective measure of disability in Parkinson's disease. Clin Rehabil 2001 Jun;15(3):241–6.
[22] Stroop J. Studies of interference in serial verbal reactions. J Exp Psychol 1935;28:643–62.
[23] Kaplan E, Goodglass H, Weintraub S. The Boston naming test. Philadelphia: Lea and Febiger; 1983.
[24] Brandt J. The Hopkins verbal learning test: development of a new memory test with six equivalent forms. Clin Neuropsychol 1991;5:125–42.
[25] Spreen O, Strauss E. A compendium of neuropsychological tests. New York: Oxford University Press; 1998.
[26] Bender R, Grouven U. Ordinal logistic regression in medical research. J R Coll Physicians Lond 1997 Sep-Oct;31(5):546–51.
[27] Chin PS, Berg AT, Spencer SS, Lee ML, Shinnar S, Sperling MR, et al. Patient-perceived impact of resective epilepsy surgery. Neurology 2006 Jun 27;66(12):1882–7.
[28] Weintraub D, Saboe K, Stern MB. Effect of age on geriatric depression scale performance in Parkinson's disease. Mov Disord 2007 Jul 15;22(9):1331–5.
[29] Karp JF, Rudy T, Weiner DK. Persistent pain biases item response on the geriatric depression scale (GDS): preliminary evidence for validity of the GDS-PAIN. Pain Med 2008 Jan–Feb;9(1):33–43.