Detecting depression in Parkinson disease

A systematic review and meta-analysis

Zahra Goodarzi, MD,
FRCPC
Kelly J. Mrklas, MSc
Derek J. Roberts, MD,
PhD
Nathalie Jette, MSc, MD,
FRCPC
Tamara Pringsheim, MSc,
MD, FRCPC
Jayna Holroyd-Leduc,
MD, FRCPC

Correspondence to
Dr. Goodarzi:
zahra.goodarzi@
albertahealthservices.ca

ABSTRACT

Background: Failure to detect depression in patients with Parkinson disease (PD) can lead to worsened outcomes for patients and caregivers. Accurate identification of depression would enable practitioners to provide comprehensive care for their patients with PD.

Methods: Our objective was to examine the diagnostic accuracy of tools for detecting depression in adult outpatients with PD. We searched MEDLINE, PsycINFO, and EMBASE (inception to December 1, 2015), gray literature, and bibliographies of included studies. The pooled prevalence of depression across studies and diagnostic accuracy estimates were calculated using random-effects models. Diagnostic accuracy estimates were calculated across the best-reported cutoffs from each study and across specific cutoffs, when feasible.

Results: Out of 8,184 citations, 21 studies were included, evaluating 24 tools, with 4 amenable to meta-analysis. The pooled prevalence of major depression was 22.9% (95% confidence interval [CI] 18.1–27.7). The 15-item Geriatric Depression Scale (GDS-15) had a pooled sensitivity of 0.81 (95% CI 0.64–0.91) and specificity of 0.91 (95% CI 0.87–0.94). The most sensitive cutoff for the GDS-15 was 5 at 0.91 (95% CI 0.83–1.00). The Beck Depression Inventory I/II had a pooled sensitivity of 0.79 (95% CI 0.61–0.90) and specificity of 0.85 (95% CI 0.79–0.90). The Montgomery-Åsberg Depression Rating Scale yielded a pooled sensitivity of 0.77 (95% CI 0.69–0.83) and specificity of 0.92 (95% CI 0.79–0.97). The Unified Parkinson's Disease Rating Scale had a pooled sensitivity of 0.72 (95% CI 0.64–0.79) and specificity of 0.80 (95% CI 0.70–0.87). All estimates had heterogeneity.

Conclusions: There are several valid tools for detecting depression in patients with PD. Practitioners should choose one that fits their clinical practice. *Neurology*® 2016;87:426–437

GLOSSARY

AUC = area under the curve; **BDI** = Beck Depression Inventory; **CI** = confidence interval; **CSDD** = Cornell Scale for Depression in Dementia; **DSM** = Diagnostic and Statistical Manual of Mental Disorders; **EDS** = Edinburgh Depression Scale; **GDS-15** = 15-item Geriatric Depression Scale; **HDRS** = Hamilton Depression Rating Scale; **ICD** = International Classification of Diseases; **IDS** = Inventory of Depressive Symptoms; **MADRS** = Montgomery-Åsberg Depression Rating Scale; **MMSE** = Mini-Mental State Examination; **NLR** = negative likelihood ratio; **PD** = Parkinson disease; **PHQ** = Patient Health Questionnaire; **PLR** = positive likelihood ratio; **QUADAS-2** = Quality Assessment of Diagnostic Accuracy Studies-2; **SROC** = summary receiver operating characteristic; **UPDRS** = Unified Parkinson's Disease Rating Scale.

Parkinson disease (PD) is a neurodegenerative disorder affecting 315 per 100,000 people in North America.¹ The prevalence of major depressive disorder is approximately 17% in those with PD, which represents a substantial proportion vs the estimated 7% in the general population in the United States² or 3.9% in Canada.³ Depression is frequently underdiagnosed and undertreated in PD, with approximately 20% of those diagnosed actually receiving treatment.^{4–6}

Comorbid depression in PD worsens cognition, function, and quality of life, and increases caregiver burden and mortality.^{4,7,8} There are several known barriers to depression therapy that impact appropriate detection and treatment. One major barrier is the symptomatic overlap between PD and depression that may mask the timely detection and appropriate treatment of depression.^{4,6,8,9}

Supplemental data
at Neurology.org

From the Departments of Community Health Sciences (Z.G., K.J.M., D.J.R., N.J., T.P., J.H.-L.) and Clinical Neurosciences (N.J., T.P.), University of Calgary; Research Priorities and Implementation (K.J.M.), Alberta Health Services, Edmonton; Departments of Critical Care Medicine (D.J.R.) and Surgery (D.J.R.), Hotchkiss Brain Institute and O'Brien Institute for Public Health (N.J., T.P., J.H.-L.), and Departments of Psychiatry and Pediatrics (T.P.) and Medicine (Z.G., J.H.-L.), University of Calgary and Alberta Health Services, Canada.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

An accurate way of detecting those patients with PD at risk for depression is highly desirable for clinicians, and would help appropriately target scarce front-line mental health resources.^{10,11} The current diagnostic gold standard for depression is the criterion-based clinical interview. However, in nonresearch settings this method is impractical for detecting depression.^{8,10} Consequently, depression screening tools are used to identify patients requiring further assessment and are a key first step to diagnosis. There are many tools available for depression screening in PD; however, it is unclear which of these tools is most accurate.^{9–12} Narrative reviews exist on this topic but a systematic review for examining tools for detecting depression in PD was not identified.^{11,13}

The objective of this study was to review the diagnostic accuracy of screening tools for the detection of depression in adult outpatients with PD. We also sought to determine whether there were any reported differences in the administration of depression screening tools.

METHODS Our study methods were determined a priori and registered with PROSPERO (CRD42014014503).¹⁴ The study protocol follows the recommendations provided in the Cochrane Handbook for Diagnostic Test Accuracy^{15,16} and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols Statement and guidelines.^{17,18}

Search strategy and literature sources. We searched MEDLINE, EMBASE, and PsycINFO from database inception to December 1, 2015, without restrictions on year or language. Controlled vocabulary was used for the 3 main clusters of PD, depression, and detection. Related keywords, MeSH, Emtree, and PsycINFO terms were combined with “or” to create search themes which were then combined with “and” (box e-1 on the *Neurology*® Web site at Neurology.org). Gray literature sources were reviewed (search date: November 24–28, 2014), including conference abstracts, society, and guideline Web sites (figure 1).

Study screening and selection. All citations and full-text articles were independently reviewed by 2 authors (Z.G., K.J.M.). If disagreement occurred, it was resolved by discussion. Eligibility criteria for the abstract review stage included the following: adults with PD and administration of a depression screening tool. Full-text eligibility criteria included diagnostic accuracy studies involving adults with PD in an outpatient setting (i.e., long-term care, assisted, or community living), the use of a depression screening tool as an index test to identify depression cases, and a gold standard. The gold standards accepted were the DSM¹⁹ or ICD. Studies were excluded if they lacked sufficient data for extraction or analysis, were unavailable in full text, or were nonhuman or not original research. We also consulted with field experts (N.J., T.P., J. Quickfall) to identify additional works. The kappa statistic was used to determine agreement between reviewers on study selection at the abstract review stage.

Data extraction. A data extraction sheet was used to record study and patient demographic variables, PD severity, dementia status, study-specific prevalence of depression, index screening test, gold or referent standard, sensitivity, specificity, predictive values, and likelihood ratios. Data were extracted and verified between 2 authors (Z.G., K.J.M.). When multiple publications were based on the same population, the study with the largest sample was chosen. Investigators contacted authors to clarify potential overlap in cases where uncertainty existed. We only included unique, nonoverlapping samples to avoid multiple publication bias.^{5,20}

Risk of bias and methodologic quality assessment. The Quality Assessment of Diagnostic Accuracy Studies– (QUADAS-2) tool²¹ was used to assess methodologic quality and risk of bias. The tool was completed in duplicate for each included study (Z.G., K.J.M.). All items within the QUADAS-2 tool were reported in 22 items after investigators collapsed 4 self-evident items together.

Data synthesis, extraction, and analyses. Two main outcomes were assessed: an estimate of pooled prevalence and the pooled diagnostic accuracy measures for depression screening tools.

Pooled prevalence. An estimate of pooled prevalence for major or major and minor depression in patients with PD was calculated from study-specific point estimates using Mantel-Haenszel-weighted DerSimonian and Laird models.²² Heterogeneity in pooled estimates was assessed using I^2 statistic and assessment of the p value associated with the Cochran Q statistic. Inspection of forest plots and the I^2 statistic were used to assess for heterogeneity. A p value of <0.05 was used to determine significance. Heterogeneity was stratified as low, moderate, substantial, or considerable (0%–40%, 30%–60%, 50%–90%, or $>75\%$, respectively).²⁰ Where heterogeneity was detected, subgroup analysis and univariate metaregression were attempted. Covariates were identified as dementia status, disease severity, clinical setting, and interview type.⁵

Pooled estimates. Individual 2×2 contingency tables were extracted and calculated from reported study-specific sensitivity, specificity, positive and negative predictive values, positive likelihood ratio (PLR) and negative likelihood ratio (NLR), and sample sizes. Zero cells were continuity corrected by adding 1.0 to each cell. Where necessary, values were back calculated from original reports and rounded to the nearest whole number for analysis. All available cutoff points were extracted for each depression screening tool. We first pooled across the optimal cutoffs reported in each study. These cutoffs may not have been the same between studies. Where possible, data were pooled across the same cutoff for a tool. When data discrepancies arose within a manuscript, authors were contacted. If data verification was not possible, those studies were excluded. This occurred in 2 studies involving 2 different tools (5-item WHO Well-Being Index and Depression Screening in Parkinson's disease questionnaire) where data inconsistencies existed between the text and tables.^{23,24}

Pooled sensitivity, specificity, and likelihood ratios were calculated using bivariate random-effects models.²⁵ The bivariate random-effects model was chosen to account for the frequent negative bivariate correlation between sensitivity and specificity. For analysis, these models require >4 studies to compute meta-analyses (midas, metandi). These bivariate models also examine the proportion of heterogeneity attributable to threshold effects. This indicates how much of the variation between studies is accounted for by differences in cutoff points. For the cutoff

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram

analyses, if only 3 studies were present, we combined estimates using DerSimonian and Laird random-effects models.

Summary receiver operating characteristic (SROC) plots were generated for all tools that were pooled to examine the changes in sensitivity and specificity values as the thresholds varied between studies.¹⁶ These plots examine test accuracy taking into account the different thresholds or cutoffs used across studies.¹⁶ More discriminatory tests generate SROC plots that are centered at the top left of the plane or with a higher area under the curve (AUC).¹⁶ To examine the posttest probabilities, a Fagan plot was created for each of the meta-analyzed tools. The calculated prevalence from included studies was used for pretest probability. Analyses were conducted using STATA 13.1 (Stata Corp., College Station, TX).

RESULTS The electronic database search strategy yielded 7,177 citations. Most studies were excluded at the screening stage because they did not examine the diagnostic accuracy of depression screening tools in patients with PD. A total of 21 studies were retained for qualitative synthesis and 15 for meta-analysis (figure 1). Screening concordance between investigators for study inclusion was good ($\kappa = 0.68$; 95% confidence interval

[CI] 0.61–0.76).²⁶ The gray literature search generated 1,007 citations (figure 1).

Risk of bias assessment. While quality scores did not lead to the removal of any studies from the review, there were concerns related to the reporting of blinding (table e-1). Blinding was generally not clearly stated; and investigators were unclear about index-referent and referent-index. The lack of clarity around blinding appeared likely attributable to unclear reporting vs design-related factors.

Summary of included studies. Overall, 21 studies including 15 unique study populations and 2,028 patients with PD were included (table 1). The mean age of participants was 67.2 years (95% CI 65.1–69.4). All included studies were in English, published between 2000 and 2014, with broad geographic representation across 10 countries (Argentina, Australia, Brazil, Ecuador, Germany, Netherlands, Sri Lanka, Turkey, United Kingdom, United States). Included diagnostic accuracy studies comprised consecutive series, direct comparison,

Table 1 Summary data of depression screening tools in outpatients with Parkinson disease (PD)

Author	Index test	No. patients with PD	Prevalence of major depression, %	Cutoff ^a	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Baillon et al., ⁴³ 2014	GDS 15	120	11.7	5/6	0.86	0.86	6.06	0.17
Chagas et al., ³⁵ 2013 ^b	GDS 15	84	22.6	>7	0.79	0.88	6.41	0.24
Dissanayaka et al., ³⁶ 2007	GDS 15	79	11.4	8/9	0.89	0.87	6.85	0.13
Thompson et al., ²⁸ 2011	GDS 15	214	14.0	>10	0.43	0.96	11.39	0.59
Tumas et al., ⁴² 2008	GDS 15	50	24.0	8/9	0.91	0.92	11.38	0.10
Weintraub et al., ³⁹ 2006	GDS 15	148	12.2	6/7	0.89	0.93	12.71	0.12
Pooled results					0.81 (0.64–0.91) ^c	0.91 (0.87–0.94) ^c	9.05 (6.46–12.67) ^c	0.21 (0.10–0.42) ^c
<i>I</i> ² (p value)					75.86 (p < 0.0001)	60.87 (p = 0.03)	0.00 (p = 0.41)	79.59 (p = 0.01)
Leentjens et al., ⁴⁴ 2000	BDI	53	23.0	13/14	0.67	0.88	5.58	0.38
Schneider et al., ²³ 2010	BDI-1a	207	15.1	14/15	0.78	0.85	5.20	0.26
Silberman et al., ⁴⁵ 2006 ^d	BDI	46	39.1	18	0.67	0.93	9.57	0.35
Tumas et al., ⁴² 2008	BDI	50	24.0	17/18	1.00	0.76	4.17	0.00
Pooled results					0.79 (0.61–0.90) ^c	0.85 (0.79–0.90) ^c	5.41 (3.83–7.63) ^c	0.25 (0.13–0.48) ^c
<i>I</i> ² (p value)					44.65 (p = 0.14)	22.67 (p = 0.27)	0.00 (p = 0.66)	0.00 (p = 0.41)
Ketharanathan et al., ³² 2014 ^e	MADRS	104	37.5	≥18	0.90	1.00	86.00	0.14
Reijnders et al., ⁴⁰ 2010	MADRS	154	17.0	13/14	0.79	0.80	3.95	0.26
Silberman et al., ⁴⁵ 2006 ^d	MADRS	46	39.1	10	0.56	0.96	14.00	0.46
Williams et al., ³³ 2012 ^f	MADRS	229	34.1	≥8	0.74	0.88	6.17	0.30
Pooled results					0.77 (0.69–0.83) ^c	0.92 (0.79–0.97) ^c	9.93 (3.36–29.38) ^c	0.25 (0.17–0.35) ^c
<i>I</i> ² (p value)					62.42 (p = 0.05)	86.99 (p < 0.0001)	65.91 (p < 0.0001)	65.08 (p = 0.04)
Starkstien and Merello, ³¹ 2007 ^f	UPDRS	168	28.5	2	0.77	0.82	4.28	0.28
Tumas et al., ⁴² 2008	UPDRS	50	24.0	0/1	0.75	0.63	2.03	0.40
Williams et al., ³³ 2012 ^f	UPDRS	229	34.1	≥1	0.70	0.77	3.04	0.39
Chagas et al., ³⁴ 2011 ^g	UPDRS	84	22.6	>2	0.61	0.92	7.14	0.43
Pooled results					0.72 (0.64–0.79) ^c	0.80 (0.70–0.87) ^c	3.60 (2.35–5.50) ^c	0.35 (0.27–0.45) ^c
<i>I</i> ² (p value)					0.00 (p = 0.65)	76.20 (p = 0.01)	26.47 (p = 0.04)	0.00 (p = 0.69)
Williams et al., ³³ 2012 ^f	GDS-30	229	34.1	≥10	0.72	0.82	4.00	0.34
Ertan et al., ⁵⁰ 2005 ^f	GDS 30	109	22.9	13/14	0.78	0.85	5.20	0.26
Weintraub et al., ³⁹ 2006 ^f	HDRS 24	148	12.2	9/10	0.88	0.78	4.00	0.15
Serrano-Dueñas et al., ³⁸ 2007	HDRS 21	115	42.0	18/19	0.86	0.95	17.73	0.14
Dissanayaka et al., ³⁶ 2007	HDRS 17	79	11.4	18/19	1.00	0.99	100.00	0.00

Continued

Table 1 Continued

Author	Index test	No. patients with PD	Prevalence of major depression, %	Cutoff ^a	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Reijnders et al., ⁴⁰ 2010 ^g	HDRS 17	224	17.0	9/10	0.98	0.75	3.92	0.03
Williams et al., ³³ 2012 ^f	HDRS 17	229	34.1	≥7	0.77	0.76	3.21	0.30
Reijnders et al., ⁴⁰ 2010	HDRS 8	224	17.0	4/5	0.98	0.71	3.38	0.03
Serrano-Dueñas et al., ³⁸ 2008	HDRS 6	115	42.0	7/8	0.79	0.91	9.57	0.22
Chagas et al., ³⁴ 2011 ^b	PHQ-2	84	22.6	>3	0.75	0.89	6.82	0.28
Chagas et al., ³⁵ 2013 ^b	PHQ-9	84	22.6	>9	1.00	0.83	5.92	0.00
Thompson et al., ²⁸ 2011	PHQ-9	214	14.0	>5	0.50	0.93	7.04	0.54
Williams et al., ³³ 2012 ^f	PHQ-9	229	34.1	≥6	0.66	0.80	3.30	0.43
Reijnders et al., ⁴⁰ 2010	MADRS modified	154	17.0	5/6	0.93	0.75	3.72	0.09
Baillon et al., ⁴³ 2014	NICE	120	11.7	0/1	1.00	0.80	5.05	0.00
Baillon et al., ⁶³ 2014 ^d	EDS Brief	120	11.7	4/5	0.93	0.66	2.73	0.11
Baillon et al., ⁶³ 2014 ^d	EDS	120	11.7	10/11	0.79	0.90	7.57	0.24
Chagas et al., ⁵¹ 2010 ^b	SDS	78	23.1	>55	0.89	0.83	5.24	0.13
Dissanayaka et al., ³⁶ 2007	HDI 17	79	11.4	15.5/16	0.89	0.93	12.71	0.12
Leentjens et al., ⁶⁹ 2001	HADS	55	22.0	18/19	0.58	0.91	6.44	0.46
Williams et al., ²⁹ 2012 ^f	BDI-II	229	34.1	≥7	0.95	0.60	2.38	0.08
Williams et al., ³³ 2009 ^f	CSDD	134	22.0	≥6	0.83	0.73	3.07	0.23
Williams et al., ³³ 2012 ^f	CESD-R	229	34.1	≥12	0.72	0.70	2.40	0.40
Williams et al., ³³ 2012 ^f	IDS-SR	229	34.1	≥14	0.90	0.60	2.25	0.17
Williams et al., ³³ 2012 ^f	IDS-C	229	34.1	≥12	0.81	0.79	3.86	0.24
Pooled from nonoverlapping studies (n = 15)		2,028 ^c	22.9 (18.1-27.7); I ² 86.7 (p < 0.001) ^c					

Abbreviations: BDI = Beck Depression Inventory; CESD-R = Center for Epidemiologic Studies Depression Scale-Revised; CI = confidence interval; CSDD = Cornell Scale for Depression in Dementia; EDS = Edinburgh Depression Scale; GDS = Geriatric Depression Scale; HADS = Hospital Anxiety and Depression Scale; HDI = Hamilton Depression Inventory; HDRS = Hamilton Depression Rating Scale; IDS-C = Inventory of Depressive Symptoms-Clinician; IDS-SR = Inventory of Depressive Symptoms-Self-report; MADRS = Montgomery-Åsberg Depression Rating Scale; NICE = National Institute for Health and Clinical Excellence; PHQ = Patient Health Questionnaire; SDS = Zung Depression Rating Scale; UPDRS = Unified Parkinson's Disease Rating Scale.

^a Cutoffs written as x/y indicate that a score of less than x indicates depression is less likely and greater than y indicates depression is more likely.

^b Prevalence for smaller sample confirmed by author.

^c Significant.

^d Age is presented as median.

^e Continuity correction of 1.

^f Sensitivity, specificity, and likelihood ratio reported for detection of combined major and minor depression. For articles by Williams et al., individual analyses without minor depression were completed and found no difference.

^g The sample in Reijnders et al.⁴⁰ overlaps with the samples in Leentjens et al.⁴¹ and Naarding et al.,³⁷ which all review MADRS and HDRS as these represent a continuous sample from a cohort. Only the largest sample⁵² was included in the table and meta-analysis.

and cross-sectional diagnostic test accuracy designs,²⁷ with the exception of 2 cohort studies.^{28,29} There were a total of 5 multicenter and 10 single-center studies (one study was unclear), and the study settings were movement disorder ($n = 11$) or neurology clinics ($n = 4$).

There were 24 validated depression screening tools assessed, including abbreviated and modified versions (table 2). Index tests were self-administered or administered by trained staff, depending on the nature of the tool. The gold standard was always administered in person by a trained professional. The most frequently used gold standard was the DSM. While data were extracted for all tools, only the 15-item Geriatric Depression Scale (GDS-15), Beck Depression Inventory (BDI) I/Ia, Unified Parkinson's Disease Rating Scale (UPDRS), and Montgomery-Åsberg Depression Rating Scale (MADRS) provided sufficient data for meta-analysis (table 1). The BDI-I and BDI-Ia were pooled, given they have similar questions and scoring. The only difference between them is the BDI-Ia collapsed multiple optional answers, each with the same score to one single answer.³⁰ Prior studies have identified it is reasonable to collapse the 2 versions.³⁰ Most studies excluded patients with dementia on the basis of a Mini-Mental State Examination (MMSE) score <23 – 24 or DSM diagnosis of dementia. There were only 3 studies that did not exclude patients with dementia^{29,31,32}; in those that did not exclude dementia, the mean or median MMSE was >24 and no studies included persons with MMSE scores below 7.

Pooled prevalence. The pooled prevalence of major depression across 15 unique populations ($n = 2,028$) was 22.9% (95% CI 18.1–27.7) (table 1). There was considerable heterogeneity between studies ($I^2 = 86.7\%$; $p < 0.001$). The prevalence estimate did not change when recalculated excluding the one study that included patients with PD with dementia.²⁹ The prevalence of major and minor depression was 31.3% (95% CI 21.1–41.6) across 8 studies and 1,201 patients. This estimate also had heterogeneity ($I^2 = 94.1\%$; $p < 0.001$). We were unable to use several demographic covariates in the final analysis due to inconsistent reporting (e.g., dementia or disease severity) or if the variable was homogeneous (e.g., clinic location). Stratified and univariate meta-regression analyses were performed on prevalence for the variables interview type and continent. These analyses were underpowered ($n < 10$, per subgroup) to detect any difference²⁷ and are thus not reported.

Tools. The estimates reported for diagnostic accuracy are for major depression, as this was consistently reported (table 2). Data from 21 studies were included in the qualitative synthesis (table 1). One of the largest studies³³ ($n = 229$) examined 9 depression

screening tools and determined the GDS-30 was the most accurate and convenient tool for use (sensitivity 72% and specificity 82%). The Patient Health Questionnaire (PHQ) is a commonly used tool, which has been validated among patients with PD within 3 separate studies (sensitivities ranging from 50% to 100%; specificities 80%–93%).^{28,33–35} Another tool was the Hamilton Depression Rating Scale (HDRS), including 24-, 21-, 17-, and 6-item versions across 6 studies.^{33,36–40} This is a commonly used clinician rating tool with reported sensitivities of 75%–100% and specificities ranging from 71% to 99%. There were 5 studies that reported accuracy estimates for the HDRS-17; however, 3 of these studies had overlapping populations,^{37,40,41} thus estimates could not be pooled. Only the largest cohort from these 3⁴⁰ is reported in the summary table. Other tools reported in single studies include National Institute for Health and Clinical Excellence, Edinburgh Depression Scale (EDS)/EDS brief, Zung Depression Rating Scale, Hospital Anxiety and Depression Scale, Beck Depression Inventory–II, Cornell Scale for Depression in Dementia (CSDD), Center for Epidemiologic Studies Depression Rating Scale–Revised, and Inventory of Depressive Symptoms (IDS)–Self-report and IDS–Clinician. The reported sensitivities for these tools vary from 58% to 100% and specificities from 60% to 93% (table 1).

The GDS-15. Six studies used the GDS-15, a self-rating tool for major depression. This tool had a pooled sensitivity of 0.81 (95% CI 0.64–0.91; $I^2 = 75.86$; $p < 0.0001$) and specificity of 0.91 (95% CI 0.87–0.94; $I^2 = 60.87$; $p = 0.03$) (figure 2A). The PLR was 9.05 (95% CI 6.46–12.67; $I^2 = 0$; $p = 0.41$) and the NLR was 0.21 (95% CI 0.10–0.42; $I^2 = 79.59$; $p = 0.01$). An SROC curve was generated to visualize the relationship between sensitivity and specificity across the 6 pooled studies with an AUC of 0.94 (95% CI 0.91–0.96). The Fagan plot for this tool found a posttest probability for confirming depression of 73% and a posttest probability of ruling out depression of 6% (figure e-1A). Only 3 studies reported estimates for use of the GDS-15 to detect major and minor depression; hence there were insufficient studies to pool.^{28,35,36,39,42,43}

For the GDS-15, cutoff points 5–9 were amenable to meta-analysis for major depression (table e-2). The cutoff of 5 had the highest sensitivity at 0.91 (95% CI 0.83–1.00; $I^2 = 0$; $p = 0.94$; $n = 3$), but the lowest specificity at 0.71 (95% CI 0.66–0.77; $I^2 = 0$; $p = 0.61$; $n = 3$). The cutoff of 6 represented a balance of sensitivity and specificity—sensitivity 0.87 (95% CI 0.78–0.95; $I^2 = 0$; $p = 0.79$) and specificity 0.88 (95% CI 0.85–0.92; $I^2 = 76.9$; $p = 0.005$). The highest specificity was at a cutoff of 9 with a specificity of 0.93% (95% CI 0.90–0.96;

Table 2 Tool characteristics for detecting depression

Tool name	Abbreviation (n = no. studies)	No. items	Rater	Permission or copyright restrictions	Estimated time to complete	Special features
Geriatric Depression Scale⁵⁵	GDS-15 (n = 6); GDS-30 (n = 2)	15; 30	Self	No	5–10 min (15-item) ⁵⁶ ; 20 min (30-item) ⁵⁷	These questions are focused on cognitive symptoms and not somatic symptoms, to avoid any overlap with normal aging or comorbid disease changes. ⁵⁰ There is no question about suicide. ²⁸
Beck Depression Inventory⁵⁸	BDI-I or Ia (n = 4); BDI-II (n = 1)	21	Self	Yes	5–10 min ⁵⁷	Versions 1 and 1A are similar in their overall design and questions, with some changing in wording to avoid double negatives. ⁵⁹ Version 2 was redesigned to fit with the DSM-IV criteria. The items cover cognitive, affective, behavioral, vegetative, and somatic symptoms. ⁵⁹
Montgomery-Åsberg Depression Rating Scale⁶⁰	MADRS (n = 4); MADRS-Mod (n = 1)	10	Clinician	Yes	15–20 min ⁵⁶	Can be used to screen, measure the severity, and specifically change in depression. ⁶⁰ Covers the DSM criteria with the exception of psychomotor retardation and agitation. ⁶⁰ Focused on psychological symptoms. ⁵⁶
Hamilton Depression Rating Scale or Inventory⁵⁴	HDRS-24 (n = 1); HDRS-21 (n = 2); HDRS-17 (n = 3); HDRS-8 (n = 1); HDRS-6 (n = 1); HDI ^a (n = 1)	24; 21; 17; 8; 6; 17	Clinician or self	No	20–30 min ⁵⁶ ; 10–15 min (for HDI) ⁵⁷	A semi-structured interview with a focus on nonsomatic symptoms. ⁵⁶ It can measure severity of depression and changes during treatment. ⁵⁴ Hamilton Depression Inventory is the self-rating version of the HDRS, available in the 17-item version. ^{36,54}
Unified Parkinson's Disease Rating Scale Part 1⁶¹	UPDRS (n = 4)	4	Self	Yes	3 min ³³	The UPDRS is a comprehensive tool for the overall assessment of PD and disability. There are 4 main sections to this tool, focused on mood, activities of daily living, motor symptoms, and complications of therapy. The first part of the tool focuses on mood and behavior, which includes one item on depression that is rated in severity from 1 to 4. ^{31,42}
Patient Health Questionnaire⁵²	PHQ 2-item (n = 1); PHQ 9-item (n = 3)	2; 9	Self	No	<5 min ⁵⁷	A scale based on the DSM-IV diagnostic criteria for depression and can monitor the severity of depressive symptoms. The PHQ-2 is the first 2 questions (depressed mood and anhedonia) of the PHQ-9 and has been demonstrated to be effective as a brief rating tool. ^{34,35}
National Institute for Health and Clinical Excellence⁵³	NICE (n = 1)	2	Self or clinician	No	Very brief	Created to screen groups at high risk of depression, with 2 items focused on the symptoms of low mood and anhedonia. ⁴³
Edinburgh Depression Scale⁶²	EDS (n = 1); EDS brief (n = 1)	10	Self	No	5 min ⁵⁷	Designed for depression in the postnatal period. However, a focus on cognitive and affective symptoms vs the somatic symptoms makes it appealing in other circumstances. ⁶³
Zung Depression Rating Scale⁶⁴	SDS (n = 1)	20	Self	No	10 min ⁵⁷	This tool is a 20-item scale, which focuses on affective, somatic, and psychological symptoms of depression. Each item is rated for severity based on frequency of symptoms. ⁵¹
Hospital Anxiety and Depression Scale⁶⁵	HADS (n = 1)	14	Self	Yes	<5 min ⁵⁷	Examines 7 items in each of depression and anxiety, which exclude somatic symptoms. For depression, the questions are focused on anhedonia. ⁵⁷
Cornell Scale for Depression in Dementia⁶⁶	CSDD (n = 1)	19	Clinician	No	30 min ⁵⁷	The CSDD is a tool designed to detect depression in patients with dementia, by including a proxy in the scales rating. ⁶⁶ This is completed after a clinician-based interview with the patient and proxy. This tool has been validated in patients with and without dementia. ^{29,66}
Center for Epidemiologic Studies Depression Rating Scale⁶⁷	CESD (n = 1)	20	Self	No	10 min ⁵⁷	Designed to assess depressive symptoms in epidemiologic studies, this tool has largely been used in research vs clinical settings. ⁵⁷ Questions were selected from other depression scales, and focus on affective and somatic symptoms. ⁶⁷
Inventory of Depressive Symptoms Clinician or Self Report⁶⁸	IDS-C (n = 1); IDS-SR (n = 1)	28	Clinician or self	No	30–45 min ⁵⁷	This tool covers the DSM-IV criteria for a major depressive episode. It also includes items on pain, anxiety, and atypical symptoms. This tool reports severity, duration, and frequency of symptoms. ⁵⁷

Abbreviations: HDI = Hamilton Depression Inventory; PD = Parkinson disease.

^a Self Rated Version.

Figure 2 Forest plots

(A-D) Forest plots of sensitivity and specificity for the Geriatric Depression Scale, Beck Depression Inventory, Montgomery-Åsberg Depression Rating Scale, and Unified Parkinson's Disease Rating Scale.

$F = 36.9$; $p = 0.21$). Overall heterogeneity was low in the cutoff analysis.

The BDI-I/IIa. Four studies used the BDI-I/IIa, a self-rating tool. The pooled sensitivity was 0.79 (95% CI 0.61–0.90; $F = 44.65$; $p = 0.14$) and specificity was 0.85 (95% CI 0.79–0.90; $F = 22.67$; $p = 0.27$) (figure 2B). The pooled PLR was 5.41 (95% CI 3.83–7.63; $F = 0$; $p = 0.66$) and NLR was 0.25 (95% CI 0.13–0.48; $F = 0$; $p = 0.41$).^{23,42,44,45} An SROC curve of sensitivity and specificity demonstrated an AUC of 0.89 (95% CI 0.86–0.92). A Fagan plot showed a posttest probability for confirming disease of 62% and the reported posttest probability of ruling out disease of 7% (figure e-1B).

The MADRS. Four studies used the MADRS, a clinician-rated tool, yielding a pooled sensitivity of 0.77 (95% CI

0.69–0.83; $F = 62.42$; $p = 0.05$) and specificity of 0.92 (95% CI 0.79–0.97; $F = 86.99$; $p < 0.0001$) (figure 2C). Pooled PLR and NLR were 9.93 (95% CI 3.36–29.38; $F = 65.91$; $p < 0.0001$) and 0.25 (95% CI 0.17–0.35; $F = 65.08$; $p = 0.04$), respectively.^{32,33,40,45} An SROC curve demonstrated an AUC of 0.84 (95% CI 0.80–0.87). The Fagan plot showed a high posttest probability for confirming disease of 75% and the reported posttest probability of ruling out disease was 7% (figure e-1C).

The UPDRS. Four studies used the UPDRS, a self-rating tool, which has a pooled sensitivity of 0.72 (95% CI 0.64–0.79; $F = 0$; $p = 0.65$) and a specificity of 0.80 (95% CI 0.70–0.87; $F = 76.20$; $p = 0.01$) (figure 2D). Pooled PLR and NLR were 3.60 (95% CI 2.35–5.50; $F = 26.47$; $p = 0.04$) and 0.35 (95% CI 0.27–0.45;

$I^2 = 0$; $p = 0.69$), respectively. An SROC curve demonstrated an AUC of 0.89 (95% CI 0.86–0.92).^{31,33,34,42} The Fagan plot showed a posttest probability for confirming disease of 52% and the reported posttest probability of ruling out disease was 9% (figure e-1D).

Heterogeneity. Statistically significant heterogeneity was observed in the pooled estimates for the GDS and MADRS tool, and could be categorized as substantial. The proportion of heterogeneity likely due to threshold effect was 1.0, indicating the observed heterogeneity was related to different cutoffs used between studies. There were not enough data to pool within the same cutoffs for the UPDRS, MADRS, or BDI. Given the high proportion of heterogeneity due to threshold effect, tools that included only 3 studies were not pooled to avoid potentially biased estimates in the setting of the different cutoffs.

DISCUSSION This systematic review identified several valid tools for detecting depression in those experiencing PD. Three tools were sensitive and specific for identifying depression in PD: the GDS-15, BDI-I/1a, and MADRS. This enables clinics to examine patients in a valid and accurate manner, and to determine which patients need further assessment. The identification of several valid depression detection tools is ideal for providers, as this allows providers to tailor their choice of tool to specific clinical circumstances.

The pooled prevalence of major depression was 22.9%, which approximates the findings from a 2008 meta-analysis that reported a prevalence of 24% for major depression in PD outpatient clinic settings ($n = 11$).⁵ It is notable that the prevalence of depression in PD varies with setting. The prevalence across all settings is 17%, and only 8% in population studies.⁵ Therefore, the use of depression detection tools is likely best targeted to outpatient clinics, given that the pretest probability for depression is higher.

Findings suggest the GDS-15 would be an appropriate screening tool for depression in this population given its higher sensitivity. Also, the GDS has the highest AUC at 0.94 on the SROC analysis, indicating better global performance. The GDS also has the advantage of being available in the public domain (copyright free) and is a well-known brief self-rating tool that requires few resources to implement in a clinic. The GDS has also been translated into many different languages and has a collateral source version.

However, there is heterogeneity in the estimate across the best-reported cutoffs, attributable to the different thresholds compared. This variation is addressed by the meta-analysis of each cutoff, which revealed a cutoff of 5 as the most robust screening cutoff with the highest sensitivity (at 91%) and nonsignificant

heterogeneity. Also, one study²⁸ in the GDS meta-analysis has a much lower reported sensitivity at 43% than the other studies, which contributed to the heterogeneity of the estimate pooled across the best cutoffs.

These findings coincide with those from a large study³³ that identified the clinical utility of the GDS. The GDS also addresses the concern of symptom overlap between depression and PD, given that the tool seeks to determine psychological and social dimensions rather than somatic dimensions, thereby reducing the potential influence of comorbid medical concerns on the outcome of the screening tool.¹³

The BDI-I/1a and MADRS are also valid tools. The BDI has a slightly lower sensitivity and PLR, making it less clinically discriminative. The main difficulty with the use of the BDI-I/1a is that it is based on the older DSM-III criteria.⁴⁶ The MADRS tool has a higher PLR, but lower sensitivity, meaning it is less able to rule out disease. Additionally, the MADRS is clinician-rated rather than self-rated, which may limit its use in certain clinical settings due to time and resource constraints. The BDI and MADRS tools are variably controlled by copyright, making these less accessible tools for use clinically.¹¹

The UPDRS is a comprehensive PD assessment tool^{34,42}; however, when the single item is isolated for depression screening, it appears to have lower discriminatory value. This is likely because a single question will not capture the varied features of depression.³³

The PHQ-9 is a commonly used depression rating tool that is based on the DSM-IV criteria, includes a question addressing suicidality, and is a self-rating tool that is in the public domain.^{33,35} However, across 3 studies, it demonstrated a wide range of sensitivity between 50% and 100%. One study reported a group of patients specifically with dementia, using the clinician CSDD.²⁹ Almost all other studies either excluded patients with dementia or included them in the total final estimate without reporting a subgroup.

Study limitations. Although we included studies from any outpatient location, the bulk of the included studies focused on movement disorders or neurology clinics. This is a potential source of bias in that patients presenting to these specialty clinics may have more severe illness, be older, be on more therapies, and potentially have a higher prevalence of comorbidities than their counterparts managed elsewhere. Additionally, there is an older age range within the included studies. It is possible that certain tools designed for an elderly population, such as the GDS,⁴⁷ may perform differently in a younger population of patients with PD. However, this was examined in 2007, and the GDS-15 demonstrated validity in nonelderly and elderly patients with PD.⁴⁸

There was a lack of reporting consistency or absence of important underlying population characteristics; for example, factors such as disease severity, patient assessment in “on” or “off” state, treatment for depression, or history of depression in the dataset.^{11,49} This meant that further subgroup analysis of the estimates based on these factors was not possible. Low study numbers for subgroup analysis precluded meaningful analysis based on factors such as interview type, location, and country.⁵

Clinical utility. The use of a tool for detecting depression would assist clinicians in identifying those patients with PD who may be at risk and require further assessment. This review demonstrates the availability of 15 validated tools for the detection of depression. Tools are potentially more practical for ease of use in the clinic if they are self-rating, without copyright, and are short, such as the GDS. However, it is important that the tool chosen provides accurate and meaningful results to the specific context to base further clinical decisions.

AUTHOR CONTRIBUTIONS

There was no direct funding for this project. Z.G. and K.J.M. performed all citation/full text screening, quality assessments, data extraction, and analysis and drafted the manuscript. K.J.M., Z.G., J.H.-L., N.J., T.P., and D.J.R. provided input and suggestions into the protocol and manuscript. Z.G. registered the protocol with PROSPERO.¹⁴ All authors reviewed and approved the final manuscript. Z.G. had full access to the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis.

ACKNOWLEDGMENT

The authors thank Diane Lorenzetti, MLS, Research Librarian, for her input into the search strategy, and William Ghali, MD, FRCPC, and Jeremy Quickfall, FRCPC, for advice and suggestions.

STUDY FUNDING

No targeted funding reported.

DISCLOSURE

Z. Goodarzi is employed as a PGY-6 Geriatric Medicine Fellow at the University of Calgary, while concurrently enrolled as a Masters' student in the Department of Community Health Sciences and cross-appointed to the Clinical Investigator Program; and has received funding from Canadian Institutes for Health Research–Canadian Graduate Student Scholarship, Alberta Innovates Health Solutions Clinician Fellowship Award, and Western Regional Training Centre affiliate award. K. Mrklas is a Doctoral student in the Department of Community Health Sciences at the University of Calgary and employed as a KT Implementation Scientist with Alberta Health Services; and is supported by an Alberta Health Services Post Secondary Education Tuition Award. D. Roberts is supported by an Alberta Innovates–Health Solutions Clinician Fellowship Award, a Knowledge Translation Canada Strategic Training in Health Research Fellowship, and funding from the Canadian Institutes of Health Research. N. Jette holds a Canada Research Chair Tier 2 in Health Services Research, is on the editorial board of *Neurology*[®], and holds research grants from the Canadian Institutes of Health Research, Alberta Innovates Health Solutions, the Alberta Spine Foundation, the University of Calgary Cumming School of Medicine, Department of Clinical Neurosciences, and Hotchkiss Brain Institute (HBI), and a grant cofunded by the University of Calgary HBI and Pfizer Canada. There is no financial support for this review. T. Pringsheim has an unrestricted educational grant from Shire Canada and research grant support from Sick Kids Foundation, Alberta Health Services Mental Health Strategic

Clinical Network, Hotchkiss Brain Institute, and Canadian Institutes for Health Research. J. Holroyd-Leduc received funding from Alberta Health Services as the Scientific Director of the Seniors Health Strategic Clinical Network. Go to Neurology.org for full disclosures.

Received December 15, 2015. Accepted in final form April 18, 2016.

REFERENCES

1. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord* 2014;29:1583–1590.
2. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005;62:617–627.
3. Patten SB, Williams JV, Lavorato DH, Wang JL, McDonald K, Bulloch AG. Descriptive epidemiology of major depressive disorder in Canada in 2012. *Can J Psychiatry* 2015;60:23–30.
4. Pachana NA, Egan SJ, Laidlaw K, et al. Clinical issues in the treatment of anxiety and depression in older adults with Parkinson's disease. *Mov Disord* 2013;28:1930–1934.
5. Reijnders JS, Ehrt U, Weber WE, Aarsland D, Leentjens AF. A systematic review of prevalence studies of depression in Parkinson's disease. *Mov Disord* 2008;23:183–189; quiz 313.
6. Frisina PG, Borod JC, Foldi NS, Tenenbaum HR. Depression in Parkinson's disease: health risks, etiology, and treatment options. *Neuropsychiatr Dis Treat* 2008;4:81–91.
7. Rocha FL, Murad MG, Stumpf BP, Hara C, Fuzikawa C. Antidepressants for depression in Parkinson's disease: systematic review and meta-analysis. *J Psychopharmacol* 2013;27:417–423.
8. Marsh L. Depression and Parkinson's disease: current knowledge. *Curr Neurol Neurosci Rep* 2013;13:409.
9. Schrag A. Apathy and depression scales in Parkinson's disease: are they good enough? *J Neurol Sci* 2011;310:216–219.
10. Schrag A, Leentjens AF. Parkinson disease: scales to detect depression in Parkinson disease. *Nat Rev Neurol* 2012;8:359–360.
11. Schrag A, Barone P, Brown RG, et al. Depression rating scales in Parkinson's disease: critique and recommendations. *Mov Disord* 2007;22:1077–1092.
12. Costa FH, Rosso AL, Maultasch H, Nicaretta DH, Vincent MB. Depression in Parkinson's disease: diagnosis and treatment. *Arq Neuropsiquiatr* 2012;70:617–620.
13. Torbey E, Pachana NA, Dissanayaka NN. Depression rating scales in Parkinson's disease: a critical review updating recent literature. *J Affect Disord* 2015;184:216–224.
14. PROSPERO: International Prospective Register of Systematic Reviews Centre for Reviews and Dissemination. Guidance Notes for Registering a Systematic Review Protocol With PROSPERO. London: National Institute for Health Research; 2013.
15. Bossuyt PM, Leflang MM. Developing criteria including studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*, version 0.4, ch 6 [updated September 2008]. London: The Cochrane Collaboration; 2008.
16. Deeks JJP, Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. London: The Cochrane Collaboration; 2013.
17. Moher D, Shamseer L, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.

18. Moher D, Liberati A, Tetzlaff J, Altman DG; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
19. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington: American Psychiatric Association Publishing; 2013.
20. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.1.0 [updated 2011]. London: The Cochrane Collaboration; 2011.
21. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–536.
22. DerSimonian R. Meta-analysis in the design and monitoring of clinical trials. *Stat Med* 1996;15:1237–1252.
23. Schneider CB, Pilhatsch M, Rifati M, et al. Utility of the WHO-Five Well-being Index as a screening tool for depression in Parkinson's disease. *Mov Disord* 2010;25:777–783.
24. Storch A, Paelecke-Habermann Y, Schneider CB, et al. Validity of the DESPAR questionnaire as a new screening tool for depression in Parkinson's disease. *Basal Ganglia* 2011;1:97–103.
25. Egger M. *Systematic Reviews in Health Care: Meta-analysis in Context*. New York: John Wiley & Sons; 2008.
26. Orwin R. Evaluating coding decisions. In: Cooper H, Hedges LV, editors. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994.
27. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. Layerthorpe: Centre for Reviews and Dissemination, University of York; 2008.
28. 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.
29. Williams JR, Marsh L. Validity of the Cornell scale for depression in dementia in Parkinson's disease with and without cognitive impairment. *Mov Disord* 2009;24:433–437.
30. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev* 1988;8:77–100.
31. 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.
32. Ketharanathan T, Hanwella R, Weerasundera R, De Silva VA. Major depressive disorder in Parkinson's disease: a cross-sectional study from Sri Lanka. *BMC Psychiatry* 2014;14:2014.
33. Williams JR, Hirsch ES, Anderson K, et al. A comparison of nine scales to detect depression in Parkinson disease: which scale to use? *Neurology* 2012;78:998–1006.
34. Chagas MH, Crippa JA, Loureiro SR, et al. Validity of the PHQ-2 for the screening of major depression in Parkinson's disease: two questions and one important answer. *Aging Ment Health* 2011;15:838–843.
35. Chagas MH, Tumas V, Rodrigues GR, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. *Age Ageing* 2013;42:645–649.
36. Dissanayaka NN, Sellbach A, Matheson S, et al. Validity of Hamilton Depression Inventory in Parkinson's disease. *Mov Disord* 2007;22:399–403.
37. Naarding P, Leentjens AF, van Kooten F, Verhey FR. Disease-specific properties of the Rating Scale for Depression in patients with stroke, Alzheimer's dementia, and Parkinson's disease. *J Neuropsychiatry Clin Neurosci* 2002;14:329–334.
38. Serrano-Duenas M, Soledad Serrano M. Concurrent validation of the 21-item and 6-item Hamilton Depression Rating Scale versus the DSM-IV diagnostic criteria to assess depression in patients with Parkinson's disease: an exploratory analysis. *Parkinsonism Relat Disord* 2008;14:233–238.
39. 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.
40. Reijnders JS, Lousberg R, Leentjens AF. Assessment of depression in Parkinson's disease: the contribution of somatic symptoms to the clinimetric performance of the Hamilton and Montgomery-Åsberg rating scales. *J Psychosom Res* 2010;68:561–565.
41. Leentjens AF, 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.
42. Tumas V, Rodrigues GG, Farias TL, Crippa JA. The accuracy of diagnosis of major depression in patients with Parkinson's disease: a comparative study among the UPDRS, the geriatric depression scale and the Beck depression inventory. *Arq Neuropsiquiatr* 2008;66:152–156.
43. Baillon S, Dennis M, Lo N, Lindesay J. Screening for depression in Parkinson's disease: the performance of two screening questions. *Age Ageing* 2014;43:200–205.
44. Leentjens AF, Verhey FR, 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.
45. 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.
46. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess* 1996;67:588–597.
47. Sheikh JI, Yesavage JA, Brooks JO III, et al. Proposed factor structure of the Geriatric Depression Scale. *Int Psychogeriatr* 1991;3:23–28.
48. Weintraub D, Saboe K, Stern MB. Effect of age on Geriatric Depression Scale performance in Parkinson's disease. *Mov Disord* 2007;22:1331–1335.
49. Marsh L, McDonald WM, Cummings J, Ravina B; NINDS/NIMH Work Group on Depression and Parkinson's Disease. Provisional diagnostic criteria for depression in Parkinson's disease: report of an NINDS/NIMH Work Group. *Mov Disord* 2006;21:148–158.
50. Ertan FS, Ertan T, Kiziltan G, Uygucil H. Reliability and validity of the Geriatric Depression Scale in depression in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2005;76:1445–1447.
51. Chagas MH, Tumas V, Loureiro SR, et al. Validity of a Brazilian version of the Zung self-rating depression scale for screening of depression in patients with Parkinson's disease. *Parkinsonism Relat Disord* 2010;16:42–45.
52. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–613.

53. National Institute for Health and Care Excellence. Depression in Adults with a Chronic Physical Health Problem: Recognition and Management. London: National Institute for Health and Care Excellence; 2009.
54. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62.
55. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1982;17:37–49.
56. Sheehan B. Assessment scales in dementia. *Ther Adv Neurol Disord* 2012;5:349–358.
57. Lam R, Michalak E, Swinson R. Assessment Scales in Depression, Mania, and Anxiety. Abingdon: Taylor Francis Group; 2005.
58. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561–571.
59. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory–II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire–9 (PHQ-9). *Arthritis Care Res* 2011;63(suppl 11):S454–S466.
60. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382–389.
61. Fahn S, Elton R. Unified Parkinson's disease rating scale. In: Fahn S, Marsden CD, Calne DB, Goldstein M, editors. *Recent Developments in Parkinson's Disease II*. Florham Park: Macmillan Health Centre Information; 1987.
62. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;150:782–786.
63. Baillon S, Lindesay J, Prabhakaran P, et al. The utility of the Edinburgh Depression Scale as a screening tool for depression in Parkinson's disease. *Int J Geriatr Psychiatry* 2014;29:1286–1293.
64. Zung WW, Richards CB, Short MJ. Self-rating depression scale in an outpatient clinic: further validation of the SDS. *Arch Gen Psychiatry* 1965;13:508–515.
65. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983;67:361–370.
66. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. *Biol Psychiatry* 1988;23:271–284.
67. Radloff LS. The Center for Epidemiological Studies–Depression Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;3:385–401.
68. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): preliminary findings. *Psychiatry Res* 1986;18:65–87.
69. Leentjens AFG, Lousberg R, Verhey FRJ. The psychometric properties of the Hospital Anxiety and Depression Scale in patients with Parkinson's disease. *Acta Neuropsychiatr* 2001;13:83–85.