Movement Disorders
CLINICAL PRACTICE
REVIEW

Outcome Measures for Parkinson's Disease Dementia: A Systematic Review

Samantha K. Holden, MD,[1] Wallace E. Jones,[1] Keith A. Baker, MD,[2] Isabel M. Boersma, MS,[1] Benzi M. Kluger, MD, MS[1,*]

Abstract: **Background:** Parkinson's disease (PD) dementia (PDD) is a major cause of morbidity and mortality in PD, which severely affects patient functioning and quality of life and increases the risk for nursing home admission. Unfortunately, current treatment options for PDD are limited and have only marginal therapeutic effects. As novel treatments are developed, there will be a need to assess their efficacy in well-designed, randomized, controlled trials. However, there is no consensus on the optimal outcome measures for use in PDD clinical trials.

Methods: A systematic review of PDD clinical trials and empirical studies of outcome measures used in PDD was performed. Outcome measures were divided into 5 categories: (1) cognitive; (2) behavioral and mood; (3) activities of daily living and quality of life; (4) global; and (5) caregiver burden.

Findings: A total of 20 PDD pharmacological clinical trials were identified. These trials incorporated a broad array of outcome measures, which were used inconsistently across trials. We summarize the psychometric properties and other relevant data on outcome measures used, including their diagnostic utility, inter-rater reliability, test-retest reliability, responsiveness, clinically meaningful change, and availability of alternate forms.

Conclusions: We have identified the best-evidenced PDD outcome measures in each domain. Further research is needed to assess the validity, reliability, and clinically meaningful change of these measures in PDD to inform the design of future clinical trials and enhance the ability of clinicians, researchers, and policy makers to interpret study results. In addition, the development of outcome measures specific to PDD may be warranted.

Parkinson's disease (PD) is the second-most common neurodegenerative illness after Alzheimer's disease (AD), affecting over 5 million persons worldwide with an anticipated doubling of prevalence in the next 20 yr.[1] PD dementia (PDD) is a major cause of morbidity and mortality in PD, affecting function, quality of life, and the risk for nursing home admission.[2] PDD has a point prevalence of approximately 31% and a mean time from onset of PD to diagnosis of dementia of 11 yr.[3] For patients surviving 20 yr or more, the chances of developing PDD may be as high as 75%.[3]

Although several effective therapies exist for PD motor symptoms, there are few treatment options available for PDD and they are of marginal clinical benefit. Currently, the only U.S. Food and Drug Administration approved treatment for PDD is the cholinesterase inhibitor, rivastigmine. A recent Cochrane review supports the use of other cholinesterase inhibitors for PDD, but acknowledges their limited clinical efficacy and the need for more-effective treatments.[4] The *N*-methyl-D-aspartate receptor antagonist, memantine, may also be beneficial in PDD; however, a large randomized, controlled trial (RCT) demonstrated its effectiveness only for dementia with Lewy bodies and not PDD.[5] Clearly, there is a need to develop and investigate more-effective therapies for PDD.

Two critical issues in designing RCTs for PDD are defining the condition and choosing appropriate outcome measures. In regard to the former, an International Parkinson and Movement Disorder Society Task Force (MDS-TF) developed clinical diagnostic criteria for PDD, with 2 levels of testing: level I, an abbreviated assessment for clinical use, and level II, more-comprehensive testing intended for research studies and

[1]Department of Neurology, University of Colorado School of Medicine, Aurora, Colorado, USA; [2]Department of Emergency Medicine, St. Luke's University Health Network, Bethlehem, Pennsylvania, USA

***Correspondence to:** Dr. Benzi M. Kluger, Department of Neurology, University of Colorado Denver, Mail Stop B-185, 12631 East 17th Avenue, Aurora, CO 80045, USA; E-mail: benzi.kluger@ucdenver.edu
Keywords: Parkinson's disease dementia, clinical trials, outcome measures.
Relevant disclosures and conflicts of interest are listed at the end of this article.
Received 15 April 2015; revised 25 June 2015; accepted 30 June 2015.
Published online 16 September 2015 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/mdc3.12225

doi:10.1002/mdc3.12225

pharmacological trials.[6,7] Since their publication in 2007, the MDS-TF diagnostic criteria have been shown to be more useful and accurate than the previous Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition standard.[8] Work continues in the application and validation of these criteria, with results indicating good agreement between MDS-TF diagnostic criteria and more-exhaustive clinical and neuropsychological testing.[9]

Concerning outcome measures, there are few empirical studies examining their performance in PDD and these are limited in terms of scales studied and psychometric properties assessed.[10] One of the more challenging aspects of assessing outcomes is defining clinically meaningful change (CMC), given that small, but statistically significant, improvements in test scores in RCTs often do not translate into any perceived benefit on the part of patients, caregivers, or clinicians.[11] An ideal outcome measure would be psychometrically valid and reliable in a PDD population, be responsive to interventions, have well-defined clinically meaningful change, and contribute to the diagnosis of PDD. Though change in cognitive test scores are clearly important, these may need to be supplemented with measures of function, quality of life, and/or caregiver burden, depending on study goals.

In this article, we review the outcome measures that have been used to date in PDD clinical trials. To evaluate the reliability and validity of the utilized measures, we summarize the data on their psychometric properties as evaluated in dementia populations. We conclude by highlighting areas where further research is needed.

Methods

Using PubMed, a literature review of clinical trials in patients with PDD from January 1965 to February 2015 was performed, using combinations of the following search terms: "Parkinson's disease dementia"; "treatment"; "intervention"; "therapy"; and "clinical trial." The outcome measures utilized in the clinical trials were then classified into 5 domains: (1) cognitive; (2) behavioral and mood; (3) activities of daily living and quality of life; (4) global; and (5) caregiver burden. For all outcome measures, we used PubMed to perform an additional search of empirical studies assessing their validity and reliability, using the search terms: "clinical significance"; "clinically meaningful change"; "validity"; "validation"; "reliability"; and "sensitivity to change." We included studies of participants with PDD, or AD, PD psychosis, and/or cognitive impairment (CI) when PDD studies were not available.

The statistical properties evaluated for each outcome measure include diagnostic utility, inter-rater reliability, test-retest reliability, responsiveness, and CMC values. *Diagnostic utility* was defined as the sensitivity and specificity of an outcome measure for diagnosis of PDD. If unavailable, concurrent validity for the measure with validated tests is reported. *Inter-rater reliability* is the degree of concordance between multiple raters, usually defined as Cohen's kappa. *Test-retest reliability* refers to the consistency of results in the same subject over a short time interval.

In a purely statistically sense, *sensitivity to change* refers to the magnitude of change that has occurred over time, as compared to a gold standard. This is slightly different than the related concept of *responsiveness*, which is the ability of a scale to detect meaningful changes over time, though the 2 terms are often used interchangeably in clinical research. To be as accurate as possible, we will use the term *responsiveness* throughout this review. Given that the true drug effect is unknown and cannot be evaluated for a single measure in isolation, we looked at the effects of the intervention for multiple measures. If the outcome of interest improved in line with other outcomes, we rated it as responsive, and if it failed to improve when other outcomes improved (or had significantly smaller effect size), we rated it as unresponsive. Responsiveness to disease progression was also looked at when data were available as a measure of responsiveness. CMC is usually defined in terms of the minimal clinically important difference, defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management."[12] The presence of alternate forms of the outcome measure and their susceptibility to practice effect are also reported, when relevant.

Based on these data, we make endorsements regarding the use of the studied outcome measures in future trials by the following criteria: (1) "recommended": the outcome measure has been applied to PDD patients and has been found valid, reliable, and responsive in psychometric studies in studies of PDD or other dementias; (2) "suggested": the outcome measure has been applied to PDD patients and psychometric studies are promising ("moderate" to "high" parameters), but not yet complete in dementia populations and/or some psychometric properties are suboptimal ("low"); and (3) "listed": the outcome measure has been applied to PDD patients, but psychometric studies have demonstrated disappointing data ("low" parameters) and/or have not yet been performed in dementia populations. These criteria were modeled after similar reviews of scales in PD and the appendix of ancillary scales to complement the MDS-UPDRS from the MDS.[13,14]

Results

A total of 20 clinical trials evaluating the efficacy of interventions in PDD were identified.[5,15–33] All of these trials assessed pharmacological treatments. Twenty-three unique outcome measures were utilized within the identified treatment trials (Tables 1 and 2).

Cognitive Outcome Measures

Cognitive scales were the most common (11 of 23) type of outcome measure identified (Table 1). The Mini–Mental State Examination (MMSE) is one of the oldest cognitive screening tools available and is the most frequently employed cognitive outcome measure in PDD trials, used in 14 studies,[22–28,31–37] despite validity concerns in PDD.[17,34] Though validated for assessing dementia severity,[10] its sensitivity and specificity for diagnosing PDD are suboptimal (80% and 74%, respectively) when tested against the most recent MDS-TF clinical diagnostic criteria.[35]

doi:10.1002/mdc3.12225

TABLE 1 Cognitive outcome measures

Outcome measure	Reference for studies used	Diagnostic utility	Inter-rater reliability	Test-retest reliability	Responsiveness	Clinically meaningful change	Endorsement
ADAS-cog	25–28, 33, 34, 36, and 37	NE	High	Moderate to high	Yes	4 points	Recommended
MDRS	23, 31, 32, and 37	High	High	High	Yes	NE	Recommended
MoCA	Not used	Moderate	High	High	Yes	NE	Suggested
CDR	25, 27, and 39	NE	NE	Moderate	Yes	NE	Suggested
FAB	33 and 34	Moderate	High	High	NE	NE	Suggested
D-KEFS VF	25–27, 34, and 36	Low to moderate	NE	Moderate to high	NE	NE	Suggested
HVLT	31	High	NE	NE	NE	NE	Suggested
TPCT	23, 27, and 33	High	NE	NE	NE	NE	Suggested
MMSE	22–28, 31–36, and 37	Low to moderate	High	High	Low for small changes	1.4 points	Listed
QTCS	22	NE	NE	NE	NE	NE	Listed
VMI	31	NE	NE	NE	NE	NE	Listed
BTA	26 and 31	NE	NE	NE	NE	NE	Listed

Endorsement definitions: (1) "Recommended": The outcome measure has been applied to PDD patients and has been found valid, reliable, and responsive in psychometric studies in studies of PDD or other dementias; (2) "Suggested": The outcome measure has been applied to PDD patients and psychometric studies are promising ("moderate" to "high" parameters), but not yet complete in dementia populations and/or some psychometric properties are suboptimal ("low"); and (3) "Listed": The outcome measure has been applied to PDD patients, but psychometric studies have demonstrated disappointing data ("low" parameters) and/or have not yet been performed in dementia populations.
QTCS, Quick Test of Cognitive Speed; VMI, Developmental Test of Visual-Motor Integration; BTA, Brief Test of Attention; PDD, Parkinson's disease dementia; NE, Not established.

Moreover, the MMSE may have floor effects in subjects with severe dementia and ceiling effects in subjects with mild cognitive impairment[38] and lacks adequate testing of executive functions.[36] In PDD, the MMSE demonstrates moderate concurrent validity with the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog; r = −0.6)[10] and low concurrent validity with the Mattis Dementia Rating Scale (MDRS; r = 0.29).[37] Though not established in PDD, the MMSE demonstrates high inter-rater reliability among cognitively impaired elderly individuals (k = 0.82–0.91)[39] and high test-retest reliability in AD (r = 0.87 and 0.89).[40,41] It lacks responsiveness to small changes, especially in patients with early PDD.[30] In the Donepezil and Memantine in Moderate to Severe Alzheimer's Disease (DOMINO) trial, CMC was defined as 1.4 points on the MMSE, based on triangulation of expert opinion and distribution-based values.[42] CMC has not been established for PDD. Although short-term practice effects have been reported in AD,[16] 2 versions of the MMSE are available for repeat testing.[43]

The ADAS-cog was utilized in 8 studies.[25–28, 33, 34, 36, 37] Though originally designed for AD, it is commonly used in PDD.[32] ADAS-cog scores demonstrate statistically significant differences between PDD patients of mild and moderate severity of CI, but its utility in diagnosis of PDD remains untested.[10] In PDD, it demonstrates moderate concurrent validity with the MMSE (r = −0.60).[10] The inter-rater reliability for PDD is not established, but it demonstrates high inter-rater reliability in AD (k = 0.91).[44] Test-retest reliability is moderate in PDD of both mild (r = 0.65) and moderate disease severity (r = 0.71).[10] The ADAS-cog is responsive in both PDD and AD, as determined by analysis of data from 3 large rivastigmine treatment trials.[45] CMC was defined as 4 points at 6 mo for the ADAS-cog in AD, though this is only meaningful for groups and not individuals.[46] Although learning effects have been reported on this exam in PDD,[30] alternate forms are available.

The MDRS[47] was used in 4 studies.[17,25,26,32] The MDRS has been validated for PDD diagnosis, with a total cut-off score of 123 yielding high sensitivity (93%) and specificity (91%).[48] However, this study has been criticized for using a sample with a low mean educational achievement (8.9 yr). A subsequent study, using subjects with a mean education of 14.8 yr, suggests this cut-off score is inadequate for detecting PDD, correctly classifying only 60.7% of the sample, with high specificity (100%), but very low sensitivity (20%).[49] This study suggests a total cut-off score of 133 instead.[49] Overall, the MDRS has been shown to be superior to the MMSE in assessing cognition[50] and, particularly, in detecting frontal lobe dysfunction in PDD.[11] Though not established for PDD, studies in AD show high inter-rater reliability (k = 0.93)[51] and high test-retest reliability (r = 0.97).[52] The MDRS is generally responsive to change in PDD trials, with the exception of 1 donepezil RCT in which it did not reveal change despite improvements in other measures.[18] CMC has not been established. An alternate form is available.[53]

The Cognitive Drug Research (CDR) system, a computer-based set of cognitive tests, was used in 3 trials.[19,21,33] Diagnostic utility of the CDR system in PDD is not established, nor is inter-rater reliability, though the computerized nature of the test mitigates inter-rater variability. For all-cause dementia, the CDR demonstrated moderate concurrent validity with the MMSE (r = 0.5–0.65).[54] The CDR-Power of Attention test demonstrated moderate test-retest reliability in PDD patients with mild dementia (r = 0.63) and moderate dementia (r = 0.46).[10] The CDR system demonstrated responsiveness in PDD pharmacological trials.[55] CMC is not established in PDD or AD. Over 20 alternate forms of the CDR are available.[56]

doi:10.1002/mdc3.12225

TABLE 2 Other outcome measures

Outcome measure	Reference for studies used	Inter-rater reliability	Test-retest reliability	Sensitivity to change	Clinically meaningful change	Endorsement
Mood/behavioral						
NPI	6, 22–27, 31–34, and 36	High	Moderate to high	Yes	8 points	Recommended
CSDD	31	High	High	Yes	NE for dementia	Recommended
BPRS	37	Moderate	NE for dementia	NE for dementia	NE for dementia	Suggested
ADLs and QOL						
ADCS-ADL	6, 25, 27, 38, and 36	High	High	Yes	2 points	Recommended
QOL-AD	7	High	High	Yes	NE for dementia	Recommended
DAD	22, 26, 33, and 34	High	High	Yes	NE for dementia	Recommended
SE-ADL	26	Moderate	NE for dementia	NE for dementia	NE for dementia	Suggested
PDQ-39	Not used	NE for dementia	Moderate	NE for dementia	Depends on subsection, but has been established	Suggested
UPDRS-ADL	31 and 37	NE for dementia	NE for dementia	NE for dementia	NE for dementia	Listed
Global						
CGI	22, 29, 35, and 37	Moderate	Moderate	Yes	NE for dementia	Recommended
ADCS-CGIC	6, 25, and 27	NE for dementia	High	Yes	1–2 points	Suggested
CIBIC+	24, 26, and 32	Low	Moderate	Yes	NE for dementia	Suggested
Caregiver burden						
ZBI	6 and 30	NE for dementia	High	Yes	13 points	Suggested

Endorsement definitions: (1) "Recommended": The outcome measure has been applied to PDD patients and has been found valid, reliable, and responsive in psychometric studies in studies of PDD or other dementias; (2) "Suggested": The outcome measure has been applied to PDD patients and psychometric studies are promising ("moderate" to "high" parameters), but not yet complete in dementia populations and/or some psychometric properties are suboptimal ("low"); and (3) "Listed": The outcome measure has been applied to PDD patients, but psychometric studies have demonstrated disappointing data ("low" parameters) and/or have not yet been performed in dementia populations. NE, Not established.

The Frontal Assessment Battery[57] (FAB) was used in 2 studies.[27,28] The FAB is not reliable as a sole test for diagnosing PDD (sensitivity 66% and specificity 72%[58]).[11] In PD and atypical parkinsonian syndromes, the FAB demonstrated high concurrent validity with the MDRS (r = 0.82) and high inter-rater reliability (k = 0.87).[57] For AD, it shows high test-retest reliability (r = 0.82) and inter-rater reliability (k = 0.98).[59] The FAB demonstrated responsiveness in both trials in which it was used. CMC is not established and there are no alternate forms.

Data on the psychometric properties of the following outcome measures for PDD are currently lacking: the Delis-Kaplan Executive Function System (D-KEFS)-Verbal Fluency Test[60] (D-KEFS VF); the Hopkins Verbal Learning Test Revised (HVLT); and the Ten Point Clock Drawing Test[61] (TPCT). The D-KEFS shows high test-retest reliability for PDD patients with mild disease severity (r = 0.79) and moderate test-retest reliability for moderate disease severity (r = 0.55).[10] The HVLT demonstrates high sensitivity (83%) and specificity (83%) for distinguishing all dementias from controls using a cut-off score of 16.[62] There are 6 alternate forms of the HVLT.[63] The TPCT shows high concurrent validity (r = 0.71–0.73) with the MMSE in AD[64] and a score of less than 8/10 identified 71% of patients with mild AD.[65] CMC has not been established for any of these measures. No data on the psychometric properties of the following outcome measures are available: A Quick Test of Cognitive Speed[66]; the Developmental Test of Visual-Motor Integration (VMI); and the Brief Test of Attention.

Although the Montreal Cognitive Assessment[67] (MoCA) has not yet been used in any trials, it is commonly used clinically and has adequate psychometric properties as a screening tool for PDD.[68] The optimal diagnostic cut-off score for the MoCA in PDD is <21/30, yielding high sensitivity (81%) and specificity (95%).[68,69] In AD and frontotemporal dementia, the MoCA shows high concurrent validity with the MMSE (r = 0.82).[70] It demonstrates high inter-rater reliability (k = 0.81) and high test-retest reliability (r = 0.79) in PD of varying cognitive involvement.[71] In a 3-yr longitudinal study of PD patients with varying levels of cognition, no significant change on the MoCA was found, even when subjects were stratified by age, MMSE score, and disease duration, suggesting that the MoCA lacks responsiveness.[72] However, in a recent trial of early AD patients, the MoCA was shown to be capable of detecting small-to-moderate cognitive change over time.[73] CMC for this measure is not established. Three alternate English versions are available.

Mood/Behavioral Outcome Measures

Mood and behavioral outcome measures are frequently employed in PDD trials (Table 2), with the Neuropsychiatric Inventory[74] (NPI) being the most commonly utilized, in 12 studies.[6,22–27,31–34,36] The NPI dysphoria subscore demonstrated moderate concurrent validity (r = 0.62) with the Hamilton Rating Scale for Depression when administered to caregivers of patients with all-cause dementia.[74] This study also showed high

doi:10.1002/mdc3.12225

inter-rater reliability for each subscale of the NPI (k = 0.89–1.0).[74] The NPI shows moderate test-retest reliability for PDD of mild (r = 0.66) and moderate severity (r = 0.73).[10] The NPI is commonly used in PD psychosis studies, with some antipsychotic treatment trials suggesting that this outcome measure may not be adequately responsive in PD populations.[75–77] However, the NPI-2, being the sum of scores for the delusions and hallucinations subscales, did reveal statistically significant changes in 1 trial of quetiapine.[75] Although CMC has not been established for PDD, the statistical analysis plan from the DOMINO trial (AD) determined that an 8-point change in the NPI was clinically meaningful, taking into account both expert opinion and distribution-based values.[42]

The Brief Psychiatric Rating Scale[78] (BPRS) was used in 1 dementia treatment trial.[32] The Brazilian version of the BPRS shows high concurrent validity (r = 0.73) of the delusion subscale and moderate concurrent validity (r = 0.43) of the hallucination subscale with the NPI.[79] Although not established for PDD, inter-rater reliability of the BPRS for AD varies significantly (range = 0.13–1.0; median = 0.45).[58] Test-retest reliability for the BPRS has not been established. Responsiveness of the BPRS has been validated in patients with mental illness[80] and PD psychosis.[75] Although not established in PDD, CMC for the BPRS in a clinical trial for PD psychosis was suggested as a change of 25%; no explanation was provided on how this value was chosen.[81]

The Cornell Scale for Depression in Dementia[82] (CSDD) was used in 1 trial.[25] The CSDD is based on observation of both the patient and informant and is validated for diagnosing depression in patients with PD of varying cognitive impairment.[83] A cut-off score of ≥8 yields sensitivity of 75% and specificity of 82% for diagnosis of depression in PD.[83] The Korean version of the CSDD in AD showed high concurrent validity with the Hamilton Depression Rating Scale (r = 0.91) and the Geriatric Depression Scale (r = 0.75).[84] Although data are not available in PDD, the CSDD has shown high inter-rater reliability in an elderly cohort with various dementias (k = 0.84)[85] and high test-retest reliability in AD (r = 0.91).[84] CMC has not been established.

Activities of Daily Living and Quality of Life Outcome Measures

We identified 5 activities of daily living (ADLs) and quality of life (QOL) measures (Table 2), with the Alzheimer's Disease Cooperative Study-Activities of Daily Living Scale (ADCS-ADL) used most frequently in 5 trials.[5,19,21,30,31] The ADCS-ADL demonstrates moderate concurrent validity with the ADAS-cog for PDD (r = −0.47).[10] Individual components of the ADCS-ADL show moderate concurrent validity with the MMSE in an AD population (r = 0.4–0.7).[86] High inter-rater reliability of the ADCS-ADL has been noted,[87] but we did not find any empiric study assessing this property. The ADCS-ADL shows high test-retest reliability for PDD patients with mild (r = 0.94) and moderate disease severity (r = 0.92).[10] The ADCS-ADL demonstrated responsiveness in detecting a treatment effect in 4 of 5 trials in which it was employed.[19,21,30,31] The ADCS-ADL is sensitive to disease progression in AD, but this property has not been established for PDD.[86] It is suggested that a change of 2 points is clinically meaningful (ie, a loss of dressing or bathing independently), but CMC has not been examined psychometrically.[87] A modified version of the ADCS-ADL, which includes a subset of 19 items, is intended for patients with moderate-to-severe dementia.

The Disability Assessment for Dementia (DAD) scale was used in 4 trials[16,20,27,28] and is a subscale of the Clinician's Interview Based Impression of Change scale (CIBIC).[27,28] In AD, the DAD shows high concurrent validity with the Rapid Disability Rating Scale-2 (r = −0.85)[88] and with the Japanese version of the Clinician Interview Based Impression of Change with caregiver input (CIBIC+) (r = 0.91).[50] The DAD demonstrates high test-retest (r = 0.96) and inter-rater reliability (k = 0.95) in AD, but has not been established for PDD.[88] The DAD demonstrated responsiveness in 2 of 4 trials in which it was utilized.[27,28] One study suggests the DAD is not sensitive to AD progression, but this property has not been assessed for PDD.[89] CMC has not been established for the DAD.

The UPDRScale-ADL Scale is not specific to dementia, but was included as an outcome measure in 2 trials.[25,32] This measure shows moderate sensitivity (67%) and specificity (79%) in discriminating PDD from PD-normal cognition when using a cutoff of 15.5.[90] The UPDRS-ADL demonstrates high inter-rater reliability (k ≥ 0.80)[91] in PD patients with moderate disease severity[8] and high test-retest reliability (r = 0.85) in early-stage PD patients,[92] though not established for PDD. The UPDRS-ADL did not demonstrate responsiveness in detecting a treatment effect in either of the 2 trials in which it was utilized. The UPDRS-ADL has shown responsiveness in disease progression for PD patients, but this property has not been evaluated in PDD.[93] One study suggests that 4 points represents CMC for the UPDRS-ADL.[92]

The Schwab and England-ADL (SE-ADL) scale was used as a secondary outcome measure in 1 trial.[20] The SE-ADL demonstrates moderate sensitivity (71%) and specificity (77%) in detecting PDD when using a cutoff of 75.[90] A study nonspecific to dementia suggests that the scale has moderate inter-rater reliability between physicians, patients, and caregivers (k = 0.6) and slightly higher inter-rater reliability between physicians and patients (k = 0.65).[55] A study of PD patients that excluded dementia suggests that the SE-ADL has moderate test-retest reliability (r = 0.70).[94] SE-ADL demonstrates responsiveness for disease progression of early PD, but this property has not been assessed specifically for PDD.[95] In the trial in which it was used, the SE-ADL did not detect a treatment effect.[20] One study suggests that a score of 12.33 represents CMC for PD patients.[94]

The Quality of Life in Alzheimer's Disease Scale (QOL-AD) was used in secondary analysis of 1 clinical trial.[15] For AD, the QOL-AD shows moderate concurrent validity with the Dementia Quality of Life scale (DQOL; r = 0.69) and the Euroqol-5D scale (r = 0.54).[96] Inter-rater reliability for PDD is not established, but the QOL-AD demonstrates moderate inter-rater reliability (k > 0.7) in AD.[96] The Chinese version of the

doi:10.1002/mdc3.12225

QOL-AD shows high test-retest reliability for patient and caregiver reports (r = 0.84 and 0.90, respectively) in an AD population.[97] The QOL-AD demonstrated responsiveness in detecting a treatment effect in the trial in which it was used.[15] CMC for the QOL-AD has not been established.

Although the 39-item Parkinson's Disease Questionnaire[98] (PDQ-39) is a commonly used outcome measure in PD, it has not been used to date in PDD trials. The PDQ-8, a short form of the PDQ-39, was used in 1 trial.[24] In a PD population, the PDQ-39 shows moderate concurrent validity with the Beck depression and anxiety inventories (r = 0.73) and the Royal Postgraduate Medical School severity scale (r = 0.66)[99] Information on inter-rater reliability of the PDQ-39 is not available for PDD. The PDQ-39 demonstrates moderate-to-high test-retest reliability in a PD population (r = 0.68–0.94),[100] and the PDQ-8 demonstrates high test-retest reliability as well (r = 0.82).[94] One study suggests that the PDQ-39 is responsive in PD patients.[99] The PDQ-39 may be insensitive to detecting change in early PD progression.[95] Minimally clinically important difference values have been determined for each subsection of the PDQ-39 in PD without dementia (mobility: −1.5, −3.2; ADL: −0.7, −4.4; emotional well-being: 0.3, −4.2; stigma: 0.8, −5.6; social support: −1.2, −11.4; cognition: 0.4, −1.8; Communication: −0.8, −4.2; bodily discomfort: 1.3, −2.1).[100]

Global Outcome Measures

Three global measures were used in the reviewed trials (Table 2). The Alzheimer's Disease Cooperative Study–Clinician's Global Impression of Change (ADCS-CGIC) was used as a primary outcome measure in 2 trials[19,21] and in another as a secondary outcome measure.[5] In AD, the ADCS-CGIC demonstrated low concurrent validity with the Global Deterioration Scale (GDS; r = 0.15) and the CDR (r = 0.15), and no significant correlation with the Functional Assessment Staging scale (FAST).[101] Inter-rater reliability of the ADCS-CGIC has not been established. The ADCS-CGIC demonstrated high short-term test-retest reliability, with 90% to 94% of subjects rated as having not changed or only minimally changed using the scale between month 1 and 2.[101] The ADCS-CGIC demonstrated responsiveness in detecting a treatment effect in 2 of the 3 trials in which it was used.[19,21] A change in score of 1 or 2 points on a 7-point scale from the ADCS-CGIC indicates CMC for AD patients.[21]

The CIBIC+, an alternate version of the CIBIC nonspecific to dementia including caregiver input, was used in 3 trials.[18,20,26] The Japanese version of the CIBIC+ demonstrated high concurrent validity in AD with the DAD (r = 0.91) and the Mental Function Impairment scale (MENFIS; r = 0.99).[50] This study also showed low inter-rater reliability (k = 0.45), but improved inter-rater reliability when collapsed into a 3-point scale (improved, no change, or worsened; k = 0.89).[102] For AD, test-retest reliability of the CIBIC+ is moderate (range = 0.4–0.6).[103] The CIBIC+ demonstrated responsiveness in detecting a treatment effect in 2 trials[18,20] and a trend toward improvement in another.[26] CMC is intrinsic to the scale, which may be used to rate patient, caregiver, and clinician impressions of change and has been validated in AD.[101]

The Clinical Global Impression scale (CGI) was used in 4 trials.[16,23,29,32] The CGI, which is not specific to dementia, consists of 3 items: the CGI-Severity scale; CGI-Improvement scale; and CGI-Efficacy index. In patients with depression and panic disorder, the CGI-Severity scale showed moderate concurrent validity (r = 0.44–0.65) and the CGI-Improvement scale showed low concurrent validity (r = 0.26–0.46) with the Hamilton Depression Rating Scale.[104] For patients with a diagnosis of dementia, the CGI-Improvement subscale demonstrates moderate inter-rater reliability (k = 0.51), whereas the inter-rater reliability of the CGI-Severity subscale was slightly higher (k = 0.66).[105] Test-retest reliability of the CGI-Severity is moderate (r = 0.65).[105] The CGI demonstrated responsiveness in detecting a statistically significant treatment effect in all 4 trials in which it was used.[16,23,29,32] CMC is intrinsic to the scale, which may be used to rate patient, caregiver, and clinicians impressions of change, but has not been validated in dementia.

Caregiver Burden

Only 1 caregiver burden outcome measure (Table 2) has been used in PDD trials to date, the Zarit Burden Interview-Caregiver Burden Assessment[106] (ZBI), which was used in 2 studies.[5,24] The ZBI demonstrates moderate-to-high concurrent validity with other measures of caregiver burden: the Burden Assessment Scale (BAS; r = 0.73) and the General Health Questionnaire (GHQ-28; r = 0.62).[107] Inter-rater reliability has not been established for the ZBI in PDD. Among caregivers of AD patients, the ZBI shows high test-retest reliability (r = 0.89).[107] Responsiveness of the ZBI in PDD is questionable, given that it detected a treatment effect in 1 trial,[24] but not in the other, though this study also did not show improvement in most cognitive test scores or the ADSC-ADL.[5] Responsiveness of the ZBI to specific interventions for caregivers of dementia patients has been demonstrated.[108] For CMC, a study of caregivers of PDD patients suggests that a ZBI score of 13 or higher in caregivers of dementia patients represents a clinically significant burden.[109]

Discussion

Twenty clinical trials evaluating the efficacy of pharmacological interventions for PDD were identified in this systematic review. Overall, the trials inconsistently employed a wide variety of outcome measures. Additionally, many of these trials use outcome measures that have not been adequately assessed for reliability or validity in PDD, though many have data in other forms of dementia. In Tables 1 and 2, we have listed endorsements of the 23 identified measures based on the available psychometric data.

In the assessment of treatments for PDD, cognitive outcome measures are of the utmost importance, both for screening of individuals for trial participation and for evaluation of the effi-

doi:10.1002/mdc3.12225

cacy of pharmacological interventions. Because almost all of these scales were developed for use in AD, it is important to pursue validation studies in PDD, particularly because the pattern of cognitive deficits observed in PDD differs from that of AD, with more executive and visuopatial dysfunction in PDD, as opposed to the more severe episodic memory impairment in AD. Cognitive measures that are biased more toward memory tasks may not be appropriate for use in PDD. The ADAS-Cog and MDRS both received "recommended" ratings. The MMSE is the most frequently utilized scale in existent PDD treatment trials, yet received a "listed" endorsement based on the available psychometric data. Further work must be done to establish inter-rater and test-retest reliability for all of these measures specifically in PDD. Future treatment trials may also consider including the MoCA in place of the MMSE, given that this scale gains traction in both clinical and research settings for PD, but further studies of responsiveness in PDD are needed.

Although cognitive outcomes have traditionally been the target of dementia treatments, noncognitive outcomes, and particularly neuropsychiatric symptoms, may have greater functional impact on patients, caregivers, and risk for nursing home placement and are considered a clinical indication for currently available agents.[110] Of the mood and behavioral outcome measures, the NPI and CSDD are the best-evidenced scales available in dementia, receiving "recommended" endorsements. An MDS-TF also designated the NPI a "recommended" scale in PD.[111]

Few ADL and QOL measures have been validated for use in PDD, though there are more data in other dementias. MDS-TF diagnostic criteria for PDD require demonstrable impairment in ADLs owing to cognitive deficits, yet there is no validated scale to assess the functional impact of cognitive impairment in PD.[6] In addition, current ADL scales may be subject to the influence of confounding variables, such as motor worsening and major depression,[90] and studies have demonstrated significant differences between patient and caregiver subjective reporting of functional disability and objective performance ratings.[112] Performance-based measures of functionality would be difficult to incorporate into treatment trials owing to equipment and time constraints, but it would be helpful to validate subjective report-based ADL scales against more objective measures. Regardless, 2 ADL scales received "recommended" endorsements, the ADCS-ADL and the DAD. QOL measures are rarely used in PDD trials, with only 1 trial employing a QOL outcome measure, that being the QOL-AD, which received a "recommended" endorsement. This is a significant gap in the field, given that the overarching goal of all disease treatment should be improvement in a patient's functionality and quality of life. The PDQ-39, which addresses both ADLs and QOL, could be a useful outcome measure in PDD treatment trials and should further study prove it valid and reliable in this population.

Global impression scales are useful in capturing minor treatment effects that may be missed by less-comprehensive measures. However, given that global measures are largely subjective, they tend to have relatively low inter-rater reliability. None of the utilized global scales have complete psychometric data in a PDD population, but the CGI does have adequate data to receive a "recommended" endorsement.

Only 2 trials incorporated a measure of caregiver burden, that being the ZBI. The ZBI is a promising measure in this domain given that it is a dementia-specific measure, though its psychometric properties have not yet been adequately evaluated, leading to a "suggested" endorsement. Regardless, we suggest that future PDD trials include the ZBI, given that caregiver burden is substantial in dementia and is an important predictor of patient institutionalization.

None of the outcome measures used in treatment trials demonstrate complete psychometric data on validity, reliability, and responsiveness in PDD, though several have complete data in dementia. Further work must be done to obtain such data in PDD. In the design of future PDD trials, we stress the importance of using dementia-specific scales that have been validated in PDD populations. Moreover, future trials should consider prioritizing outcome measures relevant to functionality and quality of life as related to cognition for PDD patients, namely, measures of cognitive functional ability. Purely cognitive measures are frequently used as primary endpoints in PDD clinical trials, yet even statistically significant improvements on such neuropsychological tests are of debatable clinical significance. As a result, many experts and even health care systems (eg, United Kingdom[113]) do not agree on the relevance of the findings of these treatment trials. From a patient-centered approach, it can be argued that the effectiveness of potential treatments for cognitive impairment should be evaluated in terms of their effect on functionality, independence, and QOL, rather than on improvements in neuropsychological testing scores. However, there is no validated diagnostic procedure currently available for establishing functional impairment related to cognition, which recent[114,115] and future studies seek to rectify.

Conclusion

Owing to the lack of empirical studies assessing the psychometric properties of dementia outcome measures, further research is needed to assess their validity and reliability in PDD populations. We propose that the outcome measures we have identified as "recommended" receive top priority for rigorous empirical validation in PDD. We also suggest that other measures commonly used in clinical practice be examined and considered for use as outcome measures in treatment trials, including the MoCA and PDQ-39. Finally, we raise the question of shifting the focus of PDD treatment trials away from neuropsychological testing scores and toward measures of functionality and QOL as related to cognition. By doing so, we could ensure that potential future treatments of PDD lead to meaningful clinical and personal improvements for patients and caregivers.

Author Roles

(1) Research Project: A. Conception, B. Organization, C. Execution; (2) Literature Search and Review: A. Design, B. Execu-

doi:10.1002/mdc3.12225

tion, C. Review and Critique; (3) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique.

S.K.H.: 1B, 1C, 2B, 2C, 3A, 3B

W.E.J.: 1C, 2B, 2C, 3B

K.A.B.: 1C, 2B, 2C, 3A, 3B

I.M.B.: 1C, 2B, 2C, 3B

B.M.K.: 1A, 1B, 1C, 2A, 2B, 2C, 3B

Disclosures

Funding Sources and Conflicts of Interest: This work was funded, in part, by the National Institutes of Health (to K.A.B.; grant no.: MH080859), the University of Colorado School of Medicine Research Track (to K.A.B.), and the National Center for Advancing Translational Sciences (NCATS; to B.M.K.; 8 KL2 TR000156-05), and National Institute of Neurological Disorders and Stroke (to B.M.K.; 1 K02 NS080885-01A1). The authors report no conflicts of interest.

Financial Disclosures for previous 12 months: Dr. Baker received funding from the University of Colorado School of Medicine Research Track, through the NIH (grant no.: MH080859). Dr. Kluger received funding from the NIH/NCATS (8 KL2 TR000156-05), U.S. Army Research Laboratory and U.S. Army Research Office (contract no.: W911NF-10-0192), the Michael J Fox Foundation, the Parkinson's Association of the Rockies (speaker), and the National Parkinson's Foundation (speaker) and served as a medical-legal expert for individual cases.

References

1. Dorsey ER, Constantinescu R, Thompson JP, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology* 2007;68:384–386.
2. Svenningsson P, Westman E, Ballard C, Aarsland D. Cognitive impairment in patients with Parkinson's disease: diagnosis, biomarkers, and treatment. *Lancet Neurol* 2012;11:697–707.
3. Hely MA, Reid WG, Adena MA, Halliday GM, Morris JG. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord* 2008;23:837–844.
4. Rolinski M, Fox C, Maidment I, McShane R. Cholinesterase inhibitors for dementia with Lewy bodies, Parkinson's disease dementia and cognitive impairment in Parkinson's disease. *Cochrane Database Syst Rev* 2012;3:CD006504.
5. Emre M, Tsolaki M, Bonuccelli U, et al. Memantine for patients with Parkinson's disease dementia or dementia with Lewy bodies: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol* 2010;9:969–977.
6. Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* 2007;22:1689–1707.
7. Dubois B, Burn D, Goetz C, et al. Diagnostic procedures for Parkinson's disease dementia: recommendations from the movement disorder society task force. *Mov Disord* 2007;22:2314–2324.
8. Martinez-Martin P, Falup-Pecurariu C, Rodriguez-Blazquez C, et al. Dementia associated with Parkinson's disease: applying the Movement Disorder Society Task Force criteria. *Parkinsonism Relat Disord* 2011;17:621–624.
9. Barton B, Grabli D, Bernard B, et al. Clinical validation of Movement Disorder Society-recommended diagnostic criteria for Parkinson's disease with dementia. *Mov Disord* 2012;27:248–253.
10. Harvey PD, Ferris SH, Cummings JL, et al. Evaluation of dementia rating scales in Parkinson's disease dementia. *Am J Alzheimers Dis Other Demen* 2010;25:142–148.
11. Courtney C, Farrell D, Gray R, et al. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet* 2004;363:2105–2115.
12. Jaeschke RS, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–415.
13. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008;23:2129–2170.
14. Friedman JH, Alves G, Hagell P, et al. Fatigue rating scales critique and recommendations by the Movement Disorders Society task force on rating scales for Parkinson's disease. *Mov Disord* 2010;25:805–822.
15. Larsson V, Engedal K, Aarsland D, Wattmo C, Minthon L, Londos E. Quality of life and the effect of memantine in dementia with lewy bodies and Parkinson's disease dementia. *Dement Geriatr Cogn Disord* 2011;32:227–234.
16. Aarsland D, Ballard C, Walker Z, et al. Memantine in patients with Parkinson's disease dementia or dementia with Lewy bodies: a double-blind, placebo-controlled, multicentre trial. *Lancet Neurol* 2009;8:613–618.
17. Aarsland D, Hutchinson M, Larsen JP. Cognitive, psychiatric and motor response to galantamine in Parkinson's disease with dementia. *Int J Geriatr Psychiatry* 2003;18:937–941.
18. Aarsland D, Laake K, Larsen JP, Janvin C. Donepezil for cognitive impairment in Parkinson's disease: a randomised controlled study. *J Neurol Neurosurg Psychiatry* 2002;72:708–712.
19. Burn D, Emre M, McKeith I, et al. Effects of rivastigmine in patients with and without visual hallucinations in dementia associated with Parkinson's disease. *Mov Disord* 2006;21:1899–1907.
20. Dubois B, Tolosa E, Katzenschlager R, et al. Donepezil in Parkinson's disease dementia: a randomized, double-blind efficacy and safety study. *Mov Disord* 2012;27:1230–1238.
21. Emre M, Aarsland D, Albanese A, et al. Rivastigmine for dementia associated with Parkinson's disease. *N Engl J Med* 2004;351:2509–2518.
22. Giladi N, Shabtai H, Gurevich T, Benbunan B, Anca M, Korczyn AD. Rivastigmine (Exelon) for dementia in patients with Parkinson's disease. *Acta Neurol Scand* 2003;108:368–373.
23. Johansson C, Ballard C, Hansson O, et al. Efficacy of memantine in PDD and DLB: an extension study including washout and open-label treatment. *Int J Geriatr Psychiatry* 2011;26:206–213.
24. Leroi I, Atkinson R, Overshott R. Memantine improves goal attainment and reduces caregiver burden in Parkinson's disease with dementia. *Int J Geriatr Psychiatry* 2014;29:899–905.
25. Leroi I, Brandt J, Reich SG, et al. Randomized placebo-controlled trial of donepezil in cognitive impairment in Parkinson's disease. *Int J Geriatr Psychiatry* 2004;19:1–8.
26. Leroi I, Overshott R, Byrne EJ, Daniel E, Burns A. Randomized controlled trial of memantine in dementia associated with Parkinson's disease. *Mov Disord* 2009;24:1217–1221.
27. Litvinenko I, Odinak M, Mogil'naya V, Emelin A. Efficacy and safety of galantamine (reminyl) for dementia in patients with Parkinson's disease (an open controlled trial). *Neurosci Behav Physiol* 2008;38:937–945.
28. Litvinenko IV, Odinak MM, Mogil'naya VI, Perstnev SV. Use of memantine (akatinol) for the correction of cognitive impairments in Parkinson's disease complicated by dementia. *Neurosci Behav Physiol* 2010;40:149–155.
29. Müller T, Welnic J, Fuchs G, Baas H, Ebersbach G, Reichmann H. The DONPAD-study—treatment of dementia in patients with Parkinson's disease with donepezil. *J Neural Transm Suppl* 2006;71:27–30.
30. Olin JT, Aarsland D, Meng X. Rivastigmine in the treatment of dementia associated with Parkinson's disease: effects on activities of daily living. *Dement Geriatr Cogn Disord* 2010;29:510–515.
31. Poewe W, Wolters E, Emre M, et al. Long-term benefits of rivastigmine in dementia associated with Parkinson's disease: an active treatment extension study. *Mov Disord* 2006;21:456–461.
32. Ravina B, Putt M, Siderowf A, et al. Donepezil for dementia in Parkinson's disease: a randomised, double blind, placebo controlled, crossover study. *J Neurol Neurosurg Psychiatry* 2005;76:934–939.
33. Wesnes KA, McKeith I, Edgar C, Emre M, Lane R. Benefits of rivastigmine on attention in dementia associated with Parkinson disease. *Neurology* 2005;65:1654–1656.

doi:10.1002/mdc3.12225

34. Mamikonyan E, Moberg PJ, Siderowf A, et al. Mild cognitive impairment is common in Parkinson's disease patients with normal Mini-Mental State Examination (MMSE) scores. *Parkinsonism Relat Disord* 2009;15:226–231.
35. Kaszas B, Kovacs N, Balas I, et al. Sensitivity and specificity of Addenbrooke's Cognitive Examination, Mattis Dementia Rating Scale, Frontal Assessment Battery and Mini Mental State Examination for diagnosing dementia in Parkinson's disease. *Parkinsonism Relat Disord* 2012;18:553–556.
36. Kulisevsky J, Pagonabarraga J. Cognitive impairment in Parkinson's disease: tools for diagnosis and assessment. *Mov Disord* 2009;24:1103–1110.
37. Freidl W, Schmidt R, Stronegger WJ, Fazekas F, Reinhart B. Sociodemographic predictors and concurrent validity of the Mini Mental State Examination and the Mattis Dementia Rating Scale. *Eur Arch Psychiatry Clin Neurosci* 1996;246:317–319.
38. Wind AW, Schellevis FG, Van Staveren G, Scholten RP, Jonker C, Van Eijk JT. Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice. *Int J Geriatr Psychiatry* 1997;12:101–108.
39. Magni E, Binetti G, Bianchetti A, Rozzini R, Trabucchi M. Mini-Mental State Examination: a normative study in Italian elderly population. *Eur J Neurol* 1996;3:198–202.
40. Clark CM, Sheppard L, Fillenbaum GG, et al. Variability in annual Mini-Mental State Examination score in patients with probable Alzheimer disease: a clinical perspective of data from the Consortium to Establish a Registry for Alzheimer's Disease. *Arch Neurol* 1999;56:857–862.
41. Fillenbaum GG, Heyman A, Wilkinson WE, Haynes CS. Comparison of two screening tests in Alzheimer's disease. The correlation and reliability of the Mini-Mental State Examination and the modified Blessed test. *Arch Neurol* 1987;44:924–927.
42. Howard R, Phillips P, Johnson T, et al. Determining the minimum clinically important differences for outcomes in the DOMINO trial. *Int J Geriatr Psychiatry* 2011;26:812–817.
43. Folstein M, Folstein SE. *Mini-Mental State Examination*. 2nd ed. Available at: http://www4.parinc.com/Products/Product.aspx?ProductID=MMSE-2 (accessed 15 January 2015).
44. Chu LW, Chiu KC, Hui SL, Yu GK, Tsui WJ, Lee PW. The reliability and validity of the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) among the elderly Chinese in Hong Kong. *Ann Acad Med Singapore* 2000;29:474–485.
45. Weintraub D, Somogyi M, Meng X. Rivastigmine in Alzheimer's disease and Parkinson's disease dementia: an ADAS-cog factor analysis. *Am J Alzheimers Dis Other Demen* 2011;26:443–449.
46. Rockwood K, Fay S, Gorman M. The ADAS-cog and clinically meaningful change in the VISTA clinical trial of galantamine for Alzheimer's disease. *Int J Geriatr Psychiatry* 2010;25:191–201.
47. Mattis S. *Dementia Rating Scale*. Odessa, FL: Psychological Assessment Resources; 1988.
48. Llebaria G, Pagonabarraga J, Kulisevsky J, et al. Cut-off score of the Mattis Dementia Rating Scale for screening dementia in Parkinson's disease. *Mov Disord* 2008;23:1546–1550.
49. Turner TH, Hinson V. Mattis Dementia Rating Scale cutoffs are inadequate for detecting dementia in Parkinson's disease. *Appl Neuropsychol Adult* 2013;20:61–65.
50. Nakamura Y, Homma A, Kobune S, et al. Reliability study on the Japanese version of the Clinician's Interview-Based Impression of Change. Analysis of subscale items and 'clinician's impression'. *Dement Geriatr Cogn Disord* 2007;23:104–115.
51. Chan IH, Siu AM. A study of the reliability and validity of the Chinese version of the Dementia Rating Scale. *Int Psychogeriatr* 2005;17:69–79.
52. Coblentz JM, Mattis S, Zingesser LH, Kasoff SS, Wisniewski HM, Katzman R. Presenile dementia. Clinical aspects and evaluation of cerebrospinal fluid dynamics. *Arch Neurol* 1973;29:299–308.
53. Schmidt KS, Lieto JM, Kiryankova E, Salvucci A. Construct and concurrent validity of the Dementia Rating Scale-2 Alternate Form. *J Clin Exp Neuropsychol* 2006;28:646–654.
54. Simpson PM, Surmon DJ, Wesnes KA, Wilcock GK. The Cognitive Drug Research computerised assessment system for demented patients: a validation study. *Int J Geriatr Psychiatry* 1991;6:95–102.
55. McRae C, Diem G, Vo A, O'Brien C, Seeberger L. Reliability of measurements of patient health status: a comparison of physician, patient, and caregiver ratings. *Parkinsonism Relat Disord* 2002;8:187–192.
56. Wesnes KA. Assessing change in cognitive function in dementia: the relative utilities of the Alzheimer's Disease Assessment Scale-Cognitive Subscale and the Cognitive Drug Research system. *Neurodegener Dis* 2008;5:261–263.
57. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. *Neurology* 2000;55:1621–1626.
58. Perrault A, Oremus M, Demers L, Vida S, Wolfson C. Review of outcome measurement instruments in Alzheimer's disease drug trials: psychometric properties of behavior and mood scales. *J Geriatr Psychiatry Neurol* 2000;13:181–196.
59. Kim TH, Huh Y, Choe JY, et al. Korean version of frontal assessment battery: psychometric properties and normative data. *Dement Geriatr Cogn Disord* 2010;29:363–370.
60. Delis DC, Kaplan E, Kramer J. *Delis Kaplan Executive Function System*. San Antionio, TX: The Psychological Corporation; 2001.
61. Watson YI, Arfken CL, Birge SJ. Clock completion: an objective screening test for dementia. *J Am Geriatr Soc* 1993;41:1235–1240.
62. Kuslansky G, Katz M, Verghese J, et al. Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. *Arch Clin Neuropsychol* 2004;19:89–104.
63. Woods SP, Scott JC, Conover E, et al. Test-retest reliability of component process variables within the Hopkins Verbal Learning Test-Revised. *Assessment* 2005;12:96–100.
64. Heinik J, Solomesh I, Berkman P. Correlation between the CAM-COG, the MMSE, and three clock drawing tests in a specialized outpatient psychogeriatric service. *Arch Gerontol Geriatr* 2004;38:77–84.
65. Manos PJ. Ten-point clock test sensitivity for Alzheimer's disease in patients with MMSE scores greater than 23. *Int J Geriatr Psychiatry* 1999;14:454–458.
66. Wiig EH, Nielsen NP, Minthon L, McPeek D, Said K, Warkentin S. Parietal lobe activation in rapid, automatized naming by adults. *Percept Mot Skills* 2002;94:1230–1244.
67. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695–699.
68. Hoops S, Nazem S, Siderowf AD, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology* 2009;73:1738–1745.
69. Dalrymple-Alford JC, MacAskill MR, Nakas CT, et al. The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology* 2010;75:1717–1725.
70. Freitas S, Simoes MR, Alves L, Santana I. Montreal Cognitive Assessment (MoCA): normative study for the Portuguese population. *J Clin Exp Neuropsychol* 2011;33:989–996.
71. Gill DJ, Freshman A, Blender JA, Ravina B. The Montreal cognitive assessment as a screening tool for cognitive impairment in Parkinson's disease. *Mov Disord* 2008;23:1043–1046.
72. Lessig S, Nie D, Xu R, Corey-Bloom J. Changes on brief cognitive instruments over time in Parkinson's disease. *Mov Disord* 2012;27:1125–1128.
73. Costa AS, Reich A, Fimm B, Ketteler ST, Schulz JB, Reetz K. Evidence of the sensitivity of the MoCA alternate forms in monitoring cognitive change in early Alzheimer's disease. *Dement Geriatr Cogn Disord* 2014;37:95–103.
74. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology* 1994;44:2308–2314.
75. Juncos JL, Roberts VJ, Evatt ML, et al. Quetiapine improves psychotic symptoms and cognition in Parkinson's disease. *Mov Disord* 2004;19:29–35.
76. Breier A, Sutton VK, Feldman PD, et al. Olanzapine in the treatment of dopamimetic-induced psychosis in patients with Parkinson's disease. *Biol Psychiatry* 2002;52:438–445.
77. Marsh L, Lyketsos C, Reich SG. Olanzapine for the treatment of psychosis in patients with Parkinson's disease and dementia. *Psychosomatics* 2001;42:477–481.

doi:10.1002/mdc3.12225

78. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep* 1962;10:799.

79. Stella F, Forlenza OV, Laks J, et al. The Brazilian version of the Neuropsychiatric Inventory-Clinician rating scale (NPI-C): reliability and validity in dementia. *Int Psychogeriatr* 2013;25:1503–1511.

80. Burlingame GM, Dunn TW, Chen S, et al. Selection of outcome assessment instruments for inpatients with severe and persistent mental illness. *Psychiatr Serv* 2005;56:444–451.

81. Friedman JH, Berman RM, Goetz CG, et al. Open-label flexible-dose pilot study to evaluate the safety and tolerability of aripiprazole in patients with psychosis associated with Parkinson's disease. *Mov Disord* 2006;21:2078–2081.

82. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. *Biol Psychiatry* 1988;23:271–284.

83. 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.

84. Lim HK, Hong SC, Won WY, Hahn C, Lee CU. Reliability and validity of the Korean version of the Cornell Scale for Depression in Dementia. *Psychiatry Investig* 2012;9:332–338.

85. Korner A, Lauritzen L, Abelskov K, et al. The Geriatric Depression Scale and the Cornell Scale for Depression in Dementia. A validity study. *Nord J Psychiatry* 2006;60:360–364.

86. Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* 1997;11 (suppl 2):S33–S39.

87. Dysken MW, Sano M, Asthana S, et al. Effect of vitamin E and memantine on functional decline in Alzheimer disease: the TEAM-AD VA cooperative randomized trial. *JAMA* 2014;311:33–44.

88. Gelinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. *Am J Occup Ther* 1999;53:471–481.

89. Trigg R, Jones RW, Knapp M, King D, Lacey LA; DADE-2 Investigator Groups. The relationship between changes in quality of life outcomes and progression of Alzheimer's disease: results from the dependence in AD in England 2 longitudinal study. *Int J Geriatr Psychiatry* 2015;30:400–408.

90. Christ JB, Fruhmann Berger M, Riedl E, et al. How precise are activities of daily living scales for the diagnosis of Parkinson's disease dementia? A pilot study *Parkinsonism Relat Disord* 2013;19:371–374.

91. Martinez-Martin P, Benito-Leon J, Alonso F, et al. Patients', doctors', and caregivers' assessment of disability using the UPDRS-ADL section: are these ratings interchangeable? *Mov Disord* 2003;18:985–992.

92. Siderowf A, McDermott M, Kieburtz K, et al. Test-retest reliability of the unified Parkinson's disease rating scale in patients with early Parkinson's disease: results from a multicenter clinical trial. *Mov Disord* 2002;17:758–763.

93. Harrison MB, Wylie SA, Frysinger RC, et al. UPDRS activity of daily living score as a marker of Parkinson's disease progression. *Mov Disord* 2009;24:224–230.

94. Dal Bello-Haas V, Klassen L, Sheppard MS, Metcalfe A. Psychometric properties of activity, self-efficacy, and quality-of-life measures in individuals with Parkinson disease. *Physiother Can* 2011;63:47–57.

95. Parashos SA, Luo S, Biglan KM, et al. Measuring disease progression in early Parkinson disease: the National Institutes of Health Exploratory Trials in Parkinson Disease (NET-PD) experience. *JAMA Neurol* 2014;71:710–716.

96. Thorgrimsen L, Selwood A, Spector A, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. *Alzheimer Dis Assoc Disord* 2003;17:201–208.

97. Yu HM, He RL, Ai YM, Liang RF, Zhou LY. Reliability and validity of the quality of life-Alzheimer disease Chinese version. *J Geriatr Psychiatry Neurol* 2013;26:230–236.

98. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Qual Life Res* 1995;4:241–248.

99. Harrison JE, Preston S, Blunt SB. Measuring symptom change in patients with Parkinson's disease. *Age Ageing* 2000;29:41–45.

100. Peto V, Jenkinson C, Fitzpatrick R. Determining minimally important differences for the PDQ-39 Parkinson's disease questionnaire. *Age Ageing* 2001;30:299–302.

101. Schneider LS, Olin JT, Doody RS, et al. Validity and reliability of the Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* 1997;11(suppl 2):S22–S32.

102. Homma A, Nakamura Y, Kobune S, et al. Reliability study on the Japanese version of the Clinician's Interview-Based Impression of Change. *Dement Geriatr Cogn Disord* 2006;21:97–103.

103. Kaduszkiewicz H, Zimmermann T, Beck-Bornholdt HP, van den Bussche H. Cholinesterase inhibitors for patients with Alzheimer's disease: systematic review of randomised clinical trials. *BMJ* 2005;331:321–327.

104. Leon AC, Shear MK, Klerman GL, Portera L, Rosenbaum JF, Goldenberg I. A comparison of symptom determinants of patient and clinician global ratings in patients with panic disorder and depression. *J Clin Psychopharmacol* 1993;13:327–331.

105. Dahlke F, Lohaus A, Gutzmann H. Reliability and clinical concepts underlying global judgments in dementia: implications for clinical research. *Psychopharmacol Bull* 1992;28:425–432.

106. Zarit S, Orr NK, Zarit JM. *The Hidden Victims of Alzheimer's Disease: Families Under Stress*. New York: New York University Press; 1985.

107. Seng BK, Luo N, Ng WY, et al. Validity and reliability of the Zarit Burden Interview in assessing caregiving burden. *Ann Acad Med Singapore* 2010;39:758–763.

108. Marim CM, Silva V, Taminato M, Barbosa DA. Effectiveness of educational programs on reducing the burden of caregivers of elderly individuals with dementia: a systematic review. [Article in English, Portuguese]. *Rev Lat Am Enfermagem* 2013;21 Spec No:267–275.

109. Gaugler JE, Mittelman MS, Hepburn K, Newcomer R. Clinically significant changes in burden and depression among dementia caregivers following nursing home admission. *BMC Med* 2010;8:85.

110. Tun SM, Murman DL, Long HL, Colenda CC, von Eye A. Predictive validity of neuropsychiatric subgroups on nursing home placement and survival in patients with Alzheimer disease. *Am J Geriatr Psychiatry* 2007;15:314–327.

111. Fernandez HH, Aarsland D, Fenelon G, et al. Scales to assess psychosis in Parkinson's disease: critique and recommendations. *Mov Disord* 2008;23:484–500.

112. Shulman LM, Pretzer-Aboff I, Anderson KE, et al. Subjective report versus objective measurement of activities of daily living in Parkinson's disease. *Mov Disord* 2006;21:794–799.

113. Getsios D, Migliaccio-Walle K, Caro JJ. NICE cost-effectiveness appraisal of cholinesterase inhibitors: was the right question posed? Were the best tools used? *Pharmacoeconomics* 2007;25:997–1006.

114. Kulisevsky J, Fernandez de Bobadilla R, Pagonabarraga J, et al. Measuring functional impact of cognitive impairment: validation of the Parkinson's disease cognitive functional rating scale. *Parkinsonism Relat Disord* 2013;19:812–817.

115. Weintraub D, Shea J, Rubright JD, et al. Psychometric properties of the brief Penn Daily Activities Questionnaire (PDAQ) for Parkinson's disease [abstract]. *Mov Disord* 2013;28:308.

doi:10.1002/mdc3.12225