HHS Public Access

Author manuscript

Mov Disord Clin Pract. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Mov Disord Clin Pract. 2016 ; 3(1): 9–18. doi:10.1002/mdc3.12225.

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], and **Benzi M. Kluger, MD, MS**[1]

[1]University of Colorado School of Medicine, Department of Neurology, Aurora, CO

[2]St. Luke's University Health Network, Department of Emergency Medicine, Bethlehem, PA

Abstract

Background—Parkinson's disease dementia (PDD) is a major cause of morbidity and mortality in Parkinson's disease (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 empiric studies of outcome measures used in PDD was performed. Outcome measures were divided into five 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 pharmacologic clinical trials were identified. These trials incorporated a broad array of outcome measures, which were used inconsistently across trials. We

*Address correspondence and reprint requests to: Dr. Benzi M. Kluger, Department of Neurology, University of Colorado Denver, Mail Stop B-185, 12631 East 17th Avenue, Aurora, CO 80045, Phone: (303) 724-2194, Fax: (303) 724-2212, benzi.kluger@ucdenver.edu.

Authorship Roles:

SKH: 1) Research project: B. Organization, C. Execution; 2) Literature search and review: B. Execution, C. Review and Critique; 3) Manuscript: A. Writing of the first draft, B. Review and Critique

WEJ: 1) Research project: C. Execution; 2) Literature search and review: B. Execution, C. Review and Critique; 3) Manuscript: B. Review and Critique

KAB: 1) Research project: C. Execution; 2) Literature search and review: B. Execution, C. Review and Critique; 3) Manuscript: A. Writing of the first draft, B. Review and Critique

IMB: 1) Research project: C. Execution; 2) Literature search and review: B. Execution, C. Review and Critique; 3) Manuscript: B. Review and Critique

BMK: 1) Research project: A. Conception, B. Organization, C. Execution; 2) Literature search and review: A. Design, C. Review and Critique; 3) Manuscript: B. Review and Critique

Conflict of Interest: There are no conflicts of interest to report.

Financial Disclosures for the Previous Twelve Months:

Dr. Holden has no financial disclosures.

Mr. Jones has no financial disclosures.

Dr. Baker received funding from the University of Colorado School of Medicine Research Track, via the National Institutes of Health (Grant # MH080859).

Ms. Boersma has no financial disclosures.

Dr. Kluger received funding from the NIH/NCATS (8 KL2 TR000156-05), US Army Research Laboratory and US Army Research Office (Contract Number W911NF-10-0192), the Michael J Fox Foundation, the Parkinson's Association of the Rockies (Speaker), the National Parkinson's Foundation (Speaker) and served as a medical-legal expert for individual cases.

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.

Keywords

Parkinson's disease dementia; clinical trials; outcome measures

Introduction

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 years.[1] Parkinson's disease 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 years.[3] For patients surviving 20 years or more, the chances of developing PDD may be as high as 75%.[3]

While 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 FDA 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 NMDA 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 (DLB) and not PDD.[5] Clearly, there is a need to develop and investigate more effective therapies for PDD.

Two critical issues in designing randomized control trials for PDD are defining the condition and choosing appropriate outcome measures. In regards to the former, a Movement Disorder Society Task Force (MDS-TF) developed clinical diagnostic criteria for PDD, with two levels of testing: level I, an abbreviated assessment for clinical use and level II, more comprehensive testing intended for research studies and 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 prior DSM-IV 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 empiric 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), as 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. While 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 paper, 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 five 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 empiric 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 clinically meaningful change 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 two terms are often used interchangeably in clinical research. To be as accurate as possible, we will use the term *responsiveness* throughout this review. As 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 was available as a measure of responsiveness. *Clinically meaningful change (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 this 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"); (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 Parkinson's disease and the appendix of ancillary scales to complement the MDS-UPDRS (Unified Parkinson Disease Rating Scale) from the MDS.[13,14]

Results

A total of twenty clinical trials evaluating the efficacy of interventions in PDD were identified.[5, 15–33] All of these trials assessed pharmacologic 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 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[22–28,31–35,37,38], despite validity concerns in PDD.[17,34] Though validated for assessing dementia severity[10], its sensitivity and specificity for diagnosing PDD are sub-optimal (80% and 74%, respectively) when tested against the most recent MDS-TF clinical diagnostic criteria.[35] Moreover, the MMSE may have floor effects in subjects with severe dementia and ceiling effects in subjects with mild cognitive impairment[36], and lacks adequate testing of executive functions.[37] In PDD, the MMSE demonstrates moderate concurrent validity with the ADAS-cog ($r=-0.6$)[10] and low concurrent validity with the MDRS ($r=0.29$).[38] 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], two versions of the MMSE are available for repeat testing.[43]

The ADAS-cog was utilized in eight studies.[25–27,28,33,34,37,38] 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 three large rivastigmine treatment trials.[45] CMC was defined as 4 points at 6 months 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 Mattis Dementia Rating Scale[47] (MDRS) was used in four studies.[17, 25, 26, 32] The MDRS has been validated for PDD diagnosis, with a total cutoff 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 years). A subsequent study, using subjects with a mean education of 14.8 years, suggests this cutoff 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 cutoff 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] While 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 one 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.[53,54]

The Cognitive Drug Research (CDR) System, a computer-based set of cognitive tests, was used in three 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).[55] 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 pharmacologic trials.[56] CMC is not established in PDD or AD. Over 20 alternate forms of the CDR are available.[57]

The Frontal Assessment Battery[58] (FAB) was used in two studies.[27, 28] The FAB is not reliable as a sole test for diagnosing PDD (sensitivity 66% and specificity 72%[59]).[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).[58] For AD, it shows high test-retest reliability (r=0.82) and inter-rater reliability (k=0.98).[60] 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 is currently lacking: the Delis-Kaplan Executive Function System-Verbal Fluency Test[61] (D-KEFS), the Hopkins Verbal Learning Test Revised (HVLT), and the Ten Point Clock Drawing Test[62] (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 cutoff score of 16.[63] There are 6 alternate forms of the HVLT.[64] The TPCT shows high concurrent validity (r=0.71–0.73) with the MMSE in AD[65] and a score of less than 8/10 identified 71% of patients with mild AD.[66] 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,[67] the Developmental Test of Visual-Motor Integration (VMI), and the Brief Test of Attention.

While the Montreal Cognitive Assessment[68] (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.[69] The optimal diagnostic cutoff score for the MoCA in PDD is < 21/30, yielding high sensitivity (81%) and specificity (95%).[69, 70] In AD and frontotemporal dementia, the MoCA shows high concurrent validity with the MMSE (r=0.82).[71] It demonstrates high inter-rater reliability (k=0.81) and high test-retest reliability (r=0.79) in PD of varying cognitive involvement.[72] In a three-year 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.[73] 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.[74] 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[75] (NPI) being the most commonly utilized, in 12 studies.[6,22–27,31–34,37] The NPI dysphoria sub-score demonstrated moderate concurrent validity (r=0.62) with the Hamilton Rating Scale for Depression when administered to caregivers of patients with all-cause dementia.[75] This study also showed high inter-rater reliability for each subscale of the NPI (k=0.89–1.0).[75] The NPI shows moderate test-retest reliability for PDD of mild severity (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.[76–78] However, the NPI-2, being the sum of scores for the delusions and hallucinations subscales, did reveal statistically significant changes in one trial of quetiapine.[76] Although CMC has not been established for PDD, the statistical analysis plan from the DOMINO trial (AD) determined that an eight point change in the NPI was clinically meaningful, taking into account both expert opinion and distribution-based values.[42]

The Brief Psychiatric Rating Scale[79] (BPRS) was used in one 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.[80] Although not established for PDD, the inter-rater reliability of the BPRS for AD varies significantly (range=0.13–1.0, median=0.45).[59] Test-retest reliability for the BPRS has not been established. The responsiveness of the BPRS has been validated in patients

with mental illness[81] and PD psychosis.[76] Although not established in PDD, the 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.[82]

The Cornell Scale for Depression in Dementia[83] (CSDD) was used in one 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.[84] A cutoff score of ≥8 yields sensitivity of 75% and specificity of 82% for diagnosis of depression in PD.[84] 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).[85] Although data is not available in PDD, the CSDD has shown high inter-rater reliability in an elderly cohort with various dementias (k=0.84)[86] and high test-retest reliability in AD (r=0.91)[85]. CMC has not been established.

Activities of Daily Living and Quality of Life Outcome Measures

We identified five activities of daily living (ADL) 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 five 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).[87] High inter-rater reliability of the ADCS-ADL has been noted,[88] 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.[87] It is suggested that a change of 2 points is clinically meaningful, i.e. a loss of dressing or bathing independently, but CMC has not been examined psychometrically.[88] 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 four 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)[89] 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.[89] 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.[90] CMC has not been established for the DAD.

The Unified Parkinson Disease Rating Scale-ADL Scale (UPDRS-ADL) is not specific to dementia but was included as an outcome measure in two 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.[91] The UPDRS-ADL demonstrates high inter-rater reliability (k≥0.80)[92] in PD patients with moderate disease severity[8] and high test-retest

reliability (r=0.85) in early stage PD patients[93], though not established for PDD. The UPDRS-ADL did not demonstrate responsiveness in detecting a treatment effect in either of the two 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.[94] One study suggests 4 points represents CMC for the UPDRS-ADL.[93]

The Schwab and England-ADL (SE-ADL) scale was used as a secondary outcome measure in one trial.[20] The SE-ADL demonstrates moderate sensitivity (71%) and specificity (77%) in detecting PDD when using a cutoff of 75.[91] A study non-specific to dementia suggests 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).[56] A study of PD patients that excluded dementia suggests the SE-ADL has moderate test-retest reliability (r=0.70).[95] SE-ADL demonstrates responsiveness for disease progression of early PD, but this property has not been assessed specifically for PDD.[96] In the trial in which it was used, the SE-ADL did not detect a treatment effect.[20] One study suggests a score of 12.33 represents CMC for PD patients.[95]

The Quality of Life in Alzheimer's Disease Scale (QOL-AD) was used in secondary analysis of one 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).[97] Inter-rater reliability for PDD is not established, but the QOL-AD demonstrates moderate inter-rater reliability (k>0.7) in AD.[97] The Chinese version of the QOL-AD shows high test-retest reliability for patient and caregiver reports (r=0.84 and 0.90 respectively) in an AD population.[98] 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 Parkinson's Disease Questionnaire[99] (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 one 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)[100] Information on the 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)[101] and the PDQ-8 demonstrates high test-retest reliability as well (r=0.82).[95] One study suggests the PDQ-39 is responsive in PD patients.[100] The PDQ-39 may be insensitive to detecting change in early PD progression.[102] 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).[101]

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 two 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 Clinical Dementia Rating Scale (CDR) (r=0.15),

and no significant correlation with the Functional Assessment Staging scale (FAST).[103] The inter-rater reliability of the ADCS-CGIC has not been established. The ADCS-CGIC demonstrated high short-term test-retest reliability, with 90–94% of subjects rated as having not changed or only minimally changed using the scale between month 1 and 2.[103] The ADCS-CGIC demonstrated responsiveness in detecting a treatment effect in two of the three 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 Clinician Interview Based Impression of Change with caregiver input (CIBIC+), an alternate version of the CIBIC non-specific to dementia including caregiver input, was used in three 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).[104] For AD, the test-retest reliability of the CIBIC+ is moderate (range=0.4–0.6).[105] The CIBIC+ demonstrated responsiveness in detecting a treatment effect in two 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 clinicians impressions of change and has been validated in AD.[103]

The Clinical Global Impression scale (CGI) was used in four trials.[16,23,29,32] The CGI, which is not specific to dementia, consists of three 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.[106] For patients with a diagnosis of dementia, the CGI-Improvement subscale demonstrates moderate inter-rater reliability ((k=0.51), while the inter-rater reliability of the CGI-Severity subscale was slightly higher (k=0.66).[107] The test-retest reliability of the CGI-Severity is moderate (r=0.65).[107] The CGI demonstrated responsiveness in detecting a statistically significant treatment effect in all four 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 one caregiver burden outcome measure (Table 2) has been used in PDD trials to date, the Zarit Burden Interview-Caregiver Burden Assessment[108] (ZBI), which was used in two 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).[109] 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).[109] The responsiveness of the ZBI in PDD is questionable, as it detected a treatment effect in one 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] The responsiveness of the ZBI to specific interventions for caregivers of dementia patients has been demonstrated.[110] 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.[111]

Discussion

Twenty clinical trials evaluating the efficacy of pharmacologic 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 Parkinson's disease dementia, cognitive outcome measures are of the utmost importance, both for screening of individuals for trial participation and for evaluation of the efficacy of pharmacologic interventions. As 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 towards 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, as this scale gains traction in both clinical and research settings for PD, but further studies of responsiveness in PDD are needed.

While cognitive outcomes have traditionally been the target of dementia treatments, non-cognitive 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.[112] 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.[113]

Few ADL and QOL measures have been validated for use in PDD, though there is more data in other dementias. The MDS-TF diagnostic criteria for PDD require demonstrable impairment in activities of daily living due 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[91], and studies have demonstrated significant differences between patient and caregiver subjective reporting of functional disability and objective performance ratings.[114] Performance-based measures of functionality would be difficult to incorporate into treatment trials due to equipment and time constraints, but it would be helpful to validate subjective report-based ADL scales against more objective measures.

Regardless, two ADL scales received “recommended” endorsements, the ADCS-ADL and the DAD. QOL measures are rarely used in PDD trials, with only one trial employing a quality of life outcome measure, being the QOL-AD, which received a “recommended” endorsement. This is a significant gap in the field, as 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, 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, since 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 two trials incorporated a measure of caregiver burden, 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, as 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 (e.g. United Kingdom[115]) 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 quality of life, 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[116, 117] and future studies seek to rectify.

Conclusion

Due to the lack of empiric 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 empiric 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 towards measures of functionality and quality of life as related to cognition. By doing so, we could ensure potential future treatments of PDD lead to meaningful clinical and personal improvements for patients and caregivers.

Acknowledgments

Funding Sources: This work was funded, in part, by the National Institutes of Health (KB: Grant #MH080859), the University of Colorado School of Medicine Research Track (KB), and the National Center for Advancing Translational Sciences (BK: 8 KL2 TR000156-05) and National Institute of Neurological Disorders and Stroke (BK: 1 K02 NS080885-01A1).

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. [PubMed: 17082464]
2. Svenningsson P, Westman E, Ballard C, Aarsland D. Cognitive impairment in patients with Parkinson's disease: diagnosis, biomarkers, and treatment. The Lancet Neurology. 2012; 11:697–707. [PubMed: 22814541]
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. Movement Disord. 2008; 23:837–844. [PubMed: 18307261]
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. [PubMed: 22419314]
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. The Lancet Neurology. 2010; 9:969–977. [PubMed: 20729148]
6. Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2007; 22:1689–1707. [PubMed: 17542011]
7. Dubois B, Burn D, Goetz C, et al. Diagnostic procedures for Parkinson's disease dementia: recommendations from the movement disorder society task force. Movement disorders : official journal of the Movement Disorder Society. 2007; 22:2314–2324. [PubMed: 18098298]
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 & related disorders. 2011; 17:621–624. [PubMed: 21684792]
9. Barton B, Grabli D, Bernard B, et al. Clinical validation of Movement Disorder Society-recommended diagnostic criteria for Parkinson's disease with dementia. Movement disorders : official journal of the Movement Disorder Society. 2012; 27:248–253. [PubMed: 22162144]
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. [PubMed: 19359706]
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. [PubMed: 15220031]
12. Jaeschke RSJ, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. Control Clin Trials. 1989; 10:407–415. [PubMed: 2691207]
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. Movement disorders : official journal of the Movement Disorder Society. 2008; 23:2129–2170. [PubMed: 19025984]
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. Movement

disorders : official journal of the Movement Disorder Society. 2010; 25:805–822. [PubMed: 20461797]

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. Dementia and geriatric cognitive disorders. 2011; 32:227–234. [PubMed: 22122992]
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. The Lancet Neurology. 2009; 8:613–618. [PubMed: 19520613]
17. Aarsland D, Hutchinson M, Larsen JP. Cognitive, psychiatric and motor response to galantamine in Parkinson's disease with dementia. International journal of geriatric psychiatry. 2003; 18:937–941. [PubMed: 14533126]
18. Aarsland D, Laake K, Larsen JP, Janvin C. Donepezil for cognitive impairment in Parkinson's disease: a randomised controlled study. Journal of neurology, neurosurgery, and 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. Movement Disord. 2006; 21:1899–1907. [PubMed: 16960863]
20. Dubois B, Tolosa E, Katzenschlager R, et al. Donepezil in Parkinson's disease dementia: a randomized, double-blind efficacy and safety study. Movement disorders : official journal of the Movement Disorder Society. 2012; 27:1230–1238. [PubMed: 22915447]
21. Emre M, Aarsland D, Albanese A, et al. Rivastigmine for dementia associated with Parkinson's disease. The New England journal of medicine. 2004; 351:2509–2518. [PubMed: 15590953]
22. Giladi N, Shabtai H, Gurevich T, Benbunan B, Anca M, Korczyn AD. Rivastigmine (Exelon) for dementia in patients with Parkinson's disease. Acta neurologica Scandinavica. 2003; 108:368–373. [PubMed: 14616309]
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. International journal of geriatric psychiatry. 2011; 26:206–213. [PubMed: 20665553]
24. Leroi I, Atkinson R, Overshott R. Memantine improves goal attainment and reduces caregiver burden in Parkinson's disease with dementia. International journal of geriatric psychiatry. 2014; 29:899–905. [PubMed: 24510471]
25. Leroi I, Brandt J, Reich SG, et al. Randomized placebo-controlled trial of donepezil in cognitive impairment in Parkinson's disease. International journal of geriatric psychiatry. 2004; 19:1–8. [PubMed: 14716693]
26. Leroi I, Overshott R, Byrne EJ, Daniel E, Burns A. Randomized controlled trial of memantine in dementia associated with Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2009; 24:1217–1221. [PubMed: 19370737]
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. [PubMed: 18975103]
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. [PubMed: 20033305]
29. Muller 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. Journal of neural transmission Supplementum. 2006:27–30. [PubMed: 17447412]
30. Olin JT, Aarsland D, Meng X. Rivastigmine in the treatment of dementia associated with Parkinson's disease: effects on activities of daily living. Dementia & Geriatric Cognitive Disorders. 2010; 29:510–515. [PubMed: 20523050]
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. Movement disorders : official journal of the Movement Disorder Society. 2006; 21:456–461. [PubMed: 16229010]

32. Ravina B, Putt M, Siderowf A, et al. Donepezil for dementia in Parkinson's disease: a randomised, double blind, placebo controlled, crossover study. Journal of neurology, neurosurgery, and 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. [PubMed: 16301500]
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 & related disorders. 2009; 15:226–231. [PubMed: 18595765]
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 & related disorders. 2012; 18:553–556. [PubMed: 22405839]
36. 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. International journal of geriatric psychiatry. 1997; 12:101–108. [PubMed: 9050431]
37. Kulisevsky J, Pagonabarraga J. Cognitive impairment in Parkinson's disease: tools for diagnosis and assessment. Movement Disord. 2009; 24:1103–1110. [PubMed: 19353727]
38. 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. European archives of psychiatry and clinical neuroscience. 1996; 246:317–319. [PubMed: 8908414]
39. Magni E, Binetti G, Bianchetti A, Rozzini R, Trabucchi M. Mini-Mental State Examination: a normative study in Italian elderly population. European journal of neurology : the official journal of the European Federation of Neurological Societies. 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. Archives of neurology. 1999; 56:857–862. [PubMed: 10404988]
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. Archives of neurology. 1987; 44:924–927. [PubMed: 3619711]
42. Howard R, Phillips P, Johnson T, et al. Determining the minimum clinically important differences for outcomes in the DOMINO trial. International journal of geriatric psychiatry. 2011; 26:812–817. [PubMed: 20848576]
43. Folstein, M.; Folstein, SE. Mini-Mental State Examination. 2[online]. Available at: http://www4.parinc.com/Products/Product.aspx?ProductID=MMSE-2
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. [PubMed: 11056778]
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. [PubMed: 22009228]
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. International journal of geriatric psychiatry. 2010; 25:191–201. [PubMed: 19548273]
47. Mattis, S. Dementia Rating Scale. Psychological Assessment Resources; Odessa, FL: 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. Movement Disord. 2008; 23:1546–1550. [PubMed: 18546326]
49. Turner TH, Hinson V. Mattis Dementia Rating Scale cutoffs are inadequate for detecting dementia in Parkinson's disease. Applied neuropsychology Adult. 2013; 20:61–65. [PubMed: 23373686]

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' Dementia and geriatric cognitive disorders. 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. International psychogeriatrics/IPA. 2005; 17:69–79. [PubMed: 15945592]
52. Coblentz JM, Mattis S, Zingesser LH, Kasoff SS, Wisniewski HM, Katzman R. Presenile dementia. Clinical aspects and evaluation of cerebrospinal fluid dynamics. Archives of neurology. 1973; 29:299–308. [PubMed: 4542729]
53. Jurica, P.; Leitten, CL.; Mattis, S. Dementia Rating Scale-2 [online]. Available at: http://www4.parinc.com/Products/Product.aspx?ProductID=DRS-2
54. Schmidt KS, Lieto JM, Kiryankova E, Salvucci A. Construct and concurrent validity of the Dementia Rating Scale-2 Alternate Form. Journal of clinical and experimental neuropsychology. 2006; 28:646–654. [PubMed: 16723314]
55. Simpson PMS, DJ, Wesnes KA, Wilcock GK. The Cognitive Drug Research computerised assessment system for demented patients: a validation study. International journal of geriatric psychiatry. 1991; 6:95–102.
56. 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 & related disorders. 2002; 8:187–192. [PubMed: 12039430]
57. 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. Neuro-degenerative diseases. 2008; 5:261–263. [PubMed: 18322407]
58. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology. 2000; 55:1621–1626. [PubMed: 11113214]
59. 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. Journal of geriatric psychiatry and neurology. 2000; 13:181–196. [PubMed: 11128058]
60. Kim TH, Huh Y, Choe JY, et al. Korean version of frontal assessment battery: psychometric properties and normative data. Dementia and geriatric cognitive disorders. 2010; 29:363–370. [PubMed: 20424455]
61. Delis, DC.; Kaplan, E.; Kramer, J. Delis Kaplan Executive Function System. San Antionio, TX: The Psychological Corporation; 2001.
62. Watson YI, Arfken CL, Birge SJ. Clock completion: an objective screening test for dementia. Journal of the American Geriatrics Society. 1993; 41:1235–1240. [PubMed: 8227899]
63. Kuslansky G, Katz M, Verghese J, et al. Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists. 2004; 19:89–104. [PubMed: 14670382]
64. 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. [PubMed: 15695747]
65. Heinik J, Solomesh I, Berkman P. Correlation between the CAMCOG, the MMSE, and three clock drawing tests in a specialized outpatient psychogeriatric service. Archives of gerontology and geriatrics. 2004; 38:77–84. [PubMed: 14599707]
66. Manos PJ. Ten-point clock test sensitivity for Alzheimer's disease in patients with MMSE scores greater than 23. International journal of geriatric psychiatry. 1999; 14:454–458. [PubMed: 10398355]
67. Wiig EH, Nielsen NP, Minthon L, McPeek D, Said K, Warkentin S. Parietal lobe activation in rapid, automatized naming by adults. Perceptual and motor skills. 2002; 94:1230–1244. [PubMed: 12186245]
68. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society. 2005; 53:695–699. [PubMed: 15817019]
69. 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. [PubMed: 19933974]

70. 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. [PubMed: 21060094]
71. Freitas S, Simoes MR, Alves L, Santana I. Montreal Cognitive Assessment (MoCA): normative study for the Portuguese population. Journal of clinical and experimental neuropsychology. 2011; 33:989–996. [PubMed: 22082082]
72. Gill DJ, Freshman A, Blender JA, Ravina B. The Montreal cognitive assessment as a screening tool for cognitive impairment in Parkinson's disease. Movement Disord. 2008; 23:1043–1046. [PubMed: 18381646]
73. Lessig S, Nie D, Xu R, Corey-Bloom J. Changes on brief cognitive instruments over time in Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2012; 27:1125–1128. [PubMed: 22692724]
74. 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. Dementia and geriatric cognitive disorders. 2014; 37:95–103. [PubMed: 24107412]
75. 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. [PubMed: 7991117]
76. Juncos JL, Roberts VJ, Evatt ML, et al. Quetiapine improves psychotic symptoms and cognition in Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2004; 19:29–35. [PubMed: 14743357]
77. 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. [PubMed: 12242060]
78. 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. [PubMed: 11815682]
79. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. Psychol Rep. 1962; 10:799.
80. 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. International psychogeriatrics/IPA. 2013; 25:1503–1511. [PubMed: 23763895]
81. Burlingame GM, Dunn TW, Chen S, et al. Selection of outcome assessment instruments for inpatients with severe and persistent mental illness. Psychiatric services. 2005; 56:444–451. [PubMed: 15812095]
82. 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. Movement disorders : official journal of the Movement Disorder Society. 2006; 21:2078–2081. [PubMed: 17013906]
83. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. Biol Psychiatry. 1988; 23:271–284. [PubMed: 3337862]
84. Williams JR, Marsh L. Validity of the Cornell scale for depression in dementia in Parkinson's disease with and without cognitive impairment. Movement disorders : official journal of the Movement Disorder Society. 2009; 24:433–437. [PubMed: 19117358]
85. 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 investigation. 2012; 9:332–338. [PubMed: 23251196]
86. 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. [PubMed: 17050293]
87. 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 disease and associated disorders. 1997; 11 (Suppl 2):S33–39. [PubMed: 9236950]
88. 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 : the journal of the American Medical Association. 2014; 311:33–44. [PubMed: 24381967]

89. Gelinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. The American journal of occupational therapy : official publication of the American Occupational Therapy Association. 1999; 53:471–481. [PubMed: 10500855]

90. Trigg R, Jones RW, Knapp M, King D, Lacey LA. 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. International journal of geriatric psychiatry. 2014

91. 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 & related disorders. 2013; 19:371–374. [PubMed: 23231974]

92. 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? Movement disorders : official journal of the Movement Disorder Society. 2003; 18:985–992. [PubMed: 14502665]

93. 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. Movement disorders : official journal of the Movement Disorder Society. 2002; 17:758–763. [PubMed: 12210871]

94. Harrison MB, Wylie SA, Frysinger RC, et al. UPDRS activity of daily living score as a marker of Parkinson's disease progression. Movement disorders : official journal of the Movement Disorder Society. 2009; 24:224–230. [PubMed: 18951537]

95. 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. Physiotherapy Canada Physiotherapie Canada. 2011; 63:47–57. [PubMed: 22210979]

96. 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 neurology. 2014; 71:710–716. [PubMed: 24711047]

97. 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 disease and associated disorders. 2003; 17:201–208. [PubMed: 14657783]

98. Yu HM, He RL, Ai YM, Liang RF, Zhou LY. Reliability and validity of the quality of life-Alzheimer disease Chinese version. Journal of geriatric psychiatry and neurology. 2013; 26:230–236. [PubMed: 23970459]

99. 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. [PubMed: 7613534]

100. Harrison JE, Preston S, Blunt SB. Measuring symptom change in patients with Parkinson's disease. Age Ageing. 2000; 29:41–45. [PubMed: 10690694]

101. Peto V, Jenkinson C, Fitzpatrick R. Determining minimally important differences for the PDQ-39 Parkinson's disease questionnaire. Age Ageing. 2001; 30:299–302. [PubMed: 11509307]

102. 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 neurology. 2014; 71:710–716. [PubMed: 24711047]

103. 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 disease and associated disorders. 1997; 11 (Suppl 2):S22–32. [PubMed: 9236949]

104. Homma A, Nakamura Y, Kobune S, et al. Reliability study on the Japanese version of the Clinician's Interview-Based Impression of Change. Dementia and geriatric cognitive disorders. 2006; 21:97–103. [PubMed: 16352896]

105. 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. [PubMed: 16081444]

106. 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. Journal of clinical psychopharmacology. 1993; 13:327–331. [PubMed: 8227491]

107. 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. [PubMed: 1296220]

108. Zarit, S.; Orr, NK.; Zarit, JM. The hidden victims of Alzheimer's disease: families under stress. New York, NY: New York University Press; 1985.

109. 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. [PubMed: 21063635]

110. 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. Revista latino-americana de enfermagem. 2013; 21(Spec No):267–275. [PubMed: 23459916]

111. 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. [PubMed: 21167022]

112. 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. The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry. 2007; 15:314–327. [PubMed: 17384314]

113. Fernandez HH, Aarsland D, Fenelon G, et al. Scales to assess psychosis in Parkinson's disease: Critique and recommendations. Movement Disord. 2008; 23:484–500. [PubMed: 18175343]

114. Shulman LM, Pretzer-Aboff I, Anderson KE, et al. Subjective report versus objective measurement of activities of daily living in Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2006; 21:794–799. [PubMed: 16482533]

115. 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. [PubMed: 18047386]

116. 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 & related disorders. 2013; 19:812–817. [PubMed: 23773412]

117. Weintraub DSJ, Rubright JD, Karlawish J, Rick J, Goldmann Gross R, Hurtig H, Duda JE, Xie SX, Siderowf A. Psychometric properties of the brief Penn daily activities questionnaire (PDAQ) for Parkinson's disease [abstract]. Movement Disord. 2013; 28:308.

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	25–27, 28, 33, 34, 37 & 38	Not Established	High	Moderate to high	Yes	4 points	Recommended
MDRS	23, 31, 32 & 38	High	High	High	Yes	Not established	Recommended
MoCA	Not used	Moderate	High	High	Yes	Not established	Suggested
CDR	25, 27 & 39	Not established	Not established	Moderate	Yes	Not established	Suggested
FAB	33 & 34	Moderate	High	High	Not established	Not established	Suggested
D-KEFS VF	25–27, 34 & 37	Low to moderate	Not established	Moderate to high	Not established	Not established	Suggested
HVLT	31	High	Not established	Not established	Not established	Not established	Suggested
TPCT	23, 27 & 33	High	Not established	Not established	Not established	Not established	Suggested
MMSE	22–28, 31–35, 37 & 38	Low to Moderate	High	High	Low for small changes	1.4 points	Listed
QTCS	22	Not established	Not established	Not established	Not established	Not established	Listed
VMI	31	Not established	Not established	Not established	Not established	Not established	Listed
BTA	26 & 31	Not established	Not established	Not established	Not established	Not established	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”)
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

Abbreviations: **MMSE**: Mini-Mental Status Examination; **ADAS-cog**: Alzheimer’s Disease Assessment Scale – Cognitive Subscale; **MDRS**: Mattis Dementia Rating Scale; **MoCA**: Montreal Cognitive Assessment; **CDR**: Cognitive Drug Research system; **FAB**: Frontal Assessment Battery; **D-KEFS VF**: Delis-Kaplan Executive Function System Verbal Fluency test; **TPCT**: Ten Point Clock Drawing Test; **QTCS**: Quick Test of Cognitive Speed; **HVLT**: Hopkins Verbal Learning Test Revised; **VMI**: Developmental Test of Visual-Motor Integration; **BTA**: Brief Test of Attention

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 & 37	High	Moderate to high	Yes	8 points	Recommended
CSDD	31	High	High	Yes	Not established for dementia	Recommended
BPRS	38	Moderate	Not established for dementia	Not established for dementia	Not established for dementia	Suggested
Activities of Daily Living and Quality of Life						
ADCS-ADL	6, 25, 27, 36 & 37	High	High	Yes	2 points	Recommended
QOL-AD	7	High	High	Yes	Not established for dementia	Recommended
DAD	22, 26, 33 & 34	High	High	Yes	Not established for dementia	Recommended
SE-ADL	26	Moderate	Not established for dementia	Not established for dementia	Not established for dementia	Suggested
PDQ-39	Not used	Not established for dementia	Moderate	Not established for dementia	Depends on subsection, but has been established	Suggested
UPDRS-ADL	31 & 38	Not established for dementia	Not established for dementia	Not established for dementia	Not established for dementia	Listed
Global						
CGI	22, 29, 35 & 38	Moderate	Moderate	Yes	Not established for dementia	Recommended
ADCS-CGIC	6, 25 & 27	Not established for dementia	High	Yes	1–2 points	Suggested
CIBIC+	24, 26 & 32	Low	Moderate	Yes	Not established for dementia	Suggested
Caregiver Burden						
ZBI	6 & 30	Not established 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”)
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

Abbreviations: **NPI**: Neuropsychiatric Inventory; **BPRS**: Brief Psychiatric Rating Scale: **CSDD**: Cornell Scale for Depression in Dementia; **ADCS-ADL**: Alzheimer’s Disease Cooperative Study-Activities of Daily Living Scale; **QOL-AD**: Quality of Life in Alzheimer’s Disease Scale; **DAD**: Disability Assessment for Dementia; **SE-ADL**: Schwab and England ADL scale; **PDQ-39**: Parkinson’s Disease Questionnaire-39; **UPDRS-ADL**: Unified Parkinson Disease Rating Scale-Activities of Daily Living; **CGI**: Clinical Global Impression scale; **ADCS-CGIC**: Alzheimer’s Disease Cooperative Study – Clinician’s Global Impression of Change; **CIBIC+**: Clinician Interview Based Impression of Change with caregiver input; **ZBI**: Zarit Burden Interview-Caregiver Burden Assessment