




OPEN ACCESS

Original research

A composite clinical motor score as a comprehensive and sensitive outcome measure for Parkinson's disease

Christine Lo ^{1,2}, Siddharth Arora,^{3,4} Michael Lawton ⁵, Thomas Barber,¹ Timothy Quinnell,⁶ Gary J Dennis,⁷ Yoav Ben-Shlomo,⁵ Michele Tao-Ming Hu ⁸

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jnnp-2021-327880>).

¹Department of Clinical Neurosciences, University of Oxford Nuffield, Oxford, UK

²Department of Clinical Neurology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

³University of Oxford Somerville College, Oxford, UK

⁴University of Oxford Said Business School, Oxford, UK

⁵Population Health Sciences, University of Bristol, Bristol, UK

⁶Papworth Hospital NHS Foundation Trust, Cambridge, UK

⁷Department of Neurology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

⁸Division of Neurology, Nuffield Department of Clinical Neurosciences, Oxford, UK

Correspondence to

Dr Christine Lo, Department of Clinical Neurosciences, University of Oxford Nuffield, Oxford OX1 1NF, Oxfordshire, UK; christine.lo@nhs.net

Received 24 August 2021

Accepted 4 March 2022

Published Online First 6 April 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Lo C, Arora S, Lawton M, et al. *J Neurol Neurosurg Psychiatry* 2022;**93**:617–624.

ABSTRACT

Background An unmet need remains for sensitive outcome measures in neuroprotective trials. The study aims to determine whether a composite clinical motor score, combining the Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) III motor examination score, Purdue Pegboard Test, and Timed Up and Go, provides greater sensitivity in detecting motor change in early disease than the MDS-UPDRS III alone.

Methods The Oxford Discovery longitudinal cohort study involves individuals with isolated rapid eye movement sleep behaviour disorder (iRBD) (n=272, confirmed polysomnographically, median follow-up: 1.6 years), idiopathic Parkinson's disease (PD) (n=909, median follow-up: 3.5 years, baseline: <3.5 years disease duration) and controls (n=316, age-matched and sex-matched, without a first-degree family history of PD). Motor and non-motor assessments were performed at each in-person visit.

Results Compared with the MDS-UPDRS III, the composite clinical motor score demonstrated a wider score distribution in iRBD and controls, lower coefficient of variation (37% vs 67%), and higher correlation coefficients with self-reported measures of motor severity (0.65 vs 0.61) and overall health status (−0.40 vs −0.33). Greater score range in mild to moderate PD, higher magnitude of longitudinal change in iRBD and longitudinal score linearity suggest better sensitivity in detecting subtle motor change. The composite clinical motor score was more accurate than the MDS-UPDRS III in predicting clinical outcomes, requiring 64% fewer participants with PD and 51% fewer participants with iRBD in sample size estimations for a hypothetical 18-month placebo-controlled clinical trial.

Conclusion The composite clinical motor score may offer greater consistency and sensitivity in detecting change than the MDS-UPDRS III.

INTRODUCTION

To date, despite multiple neuroprotective treatment trials, none has been definitively proven to alter the progression of Parkinson's disease (PD).¹ Their failure, at least in part, has been blamed on the use of insensitive endpoints and outcome measures.^{2,3}

The Purdue Pegboard Test,⁴ a test of manual dexterity with established use across disease and age groups, has been associated with high test–retest reliability (intraclass correlation coefficient (ICC)

Key messages

What is already known on this topic?

⇒ There remains an unmet need for sensitive outcome measures in Parkinson's disease (PD) and isolated rapid eye movement sleep behaviour disorder (iRBD).

What this study adds?

⇒ We present a composite clinical motor score and demonstrate its consistency and sensitivity in detecting change, resulting in 64% reduction in participants with PD and 51% reduction in participants with iRBD required for a hypothetical 18-month placebo-controlled clinical trial.

How this study might affect research, practice or policy

⇒ The composite clinical motor score holds promise for use in future neuroprotective trials where it may assist in detecting subtle motor change while permitting a reduction in sample size.

≥0.90).⁵ It has been used to distinguish individuals with PD from controls, with an area under the curve (AUC=0.8) on par with that of movement disorder ratings of the respective Unified Parkinson's Disease Rating Scale (UPDRS) III item⁶ and in predicting incident PD (HR 1.35, 95% CI 1.11 to 1.67) in a large population-based study.⁷

Another semiquantitative test of motor function, the Timed Up and Go (TUG),⁸ has been used extensively in Parkinson's, demonstrating good to excellent test–retest and intrarater and inter-rater reliability, with ICC estimates consistently in the range of 0.87–0.99.^{9–11} It is the most frequently used test in the assessment of functional mobility in PD and has been used to assess the efficacy of therapeutic interventions.^{12,13} There is some evidence to suggest that it may have greater sensitivity to detect change than the UPDRS III. A study that evaluated the long-term effect of bilateral subthalamic nucleus deep brain stimulation identified a greater improvement in TUG scores with stimulation, compared with UPDRS III assessed change in gait (70% vs 56% reduction in scores, respectively).¹³ Similarly, in a randomised controlled trial assessing the effect of intensive goal-based rehabilitation in addition

to rotigotine, versus rotigotine alone, no differences in UPDRS total, II or III scores were identified, in contrast to an improvement in the TUG ($p=0.03$).¹²

Nonetheless, the original total UPDRS,¹⁴ superseded by the Movement Disorders Society-sponsored revision of the UPDRS (MDS-UPDRS),¹⁵ and parts thereof, remains the most popular outcome measures in clinical trials. Of MDS-UPDRS parts I–III, it is the part III motor examination score that is associated with the highest test–retest (intrarater) reliability.¹⁶ Concordance between those rating the motor examination varies according to the level of experience¹⁷ and the individual item being rated¹⁸; differences between raters can surpass the 3.25-point minimum change associated with a clinically meaningful difference to patients.^{17–19} Significant within-subject measurement error has been noted when using the MDS-UPDRS III to track changes in PD progression over time, with estimates of within-subject reliability being lowest at 0.23 (95% CI 0.10 to 0.43) in the ON state.²⁰ Yet, despite the known weaknesses of the MDS-UPDRS III, in the absence of an alternative widely accepted measure of PD motor severity, its popularity persists.

One of the critiques levelled at the MDS-UPDRS III in the quantification of motor severity is its floor effect and insensitivity when measuring early disease.²¹ Abnormalities in semi-quantitative motor testing have been found to predate and more strongly predict phenoconversion in individuals with isolated rapid eye movement sleep behaviour disorder (iRBD) compared with abnormalities detected on UPDRS/MDS-UPDRS III motor examination (HR: 3.16 vs 3.03).²² Recognising the importance of ancillary outcome measures, the TUG and Purdue Pegboard Test are being used alongside the UPDRS as secondary outcome measures in the phase II evaluation of amroxol for treatment of PD dementia.²³

Hypothesis

We hypothesised that a composite clinical motor score, combining the MDS-UPDRS III motor examination score, the Purdue Pegboard Test and the TUG, would provide a more comprehensive overview of individual participant symptom severity and allow greater sensitivity in detecting motor change in early disease than the MDS-UPDRS III alone.

To the best of our knowledge, this is the first study that undertakes a detailed evaluation of a composite clinical motor score in participants with iRBD and early-stage PD combined.

MATERIALS AND METHODS

Study participants

Data described pertain to participants within the previously detailed UK Oxford Discovery study,^{24–25} a longitudinal cohort study involving individuals with (1) iRBD (polysomnographically confirmed in keeping with International Classification of Sleep Disorders Third Edition criteria),²⁶ recruited from three sleep centres: John Radcliffe Hospital, Oxford; Royal Papworth Hospital, Cambridge; and Royal Hallamshire Hospital, Sheffield; (2) idiopathic PD (fulfilling the UK PD Brain Bank criteria for probable PD²⁷ and of less than 3.5 years disease duration from diagnosis at the point of recruitment) recruited from neurology clinics across the Thames Valley area; and (3) age-matched and sex-matched controls (without a first-degree family history of PD), the spouses and friends of participants with PD. Other obligatory inclusion criteria included age >18 years old, fluency in English and the absence of cognitive impairment/dementia that would preclude the provision of informed

consent. Recruitment was contingent on the provision of written informed consent.

Clinical assessments

All participants were reviewed in person in clinic at baseline. Fifty controls were invited to return for a further single in-person assessment, with the rest of the control visits carried out remotely via telephone. Participants with iRBD and PD were preferentially seen in clinic at their 18 monthly assessments; telephone assessments were carried out for participants unwilling or unable to attend in person. Data from telephone assessments were excluded from analyses given the absence of clinical motor assessments.

The ensuing clinical motor assessments were performed exclusively in person and included (1) the MDS-UPDRS III motor examination¹⁵; (2) the Purdue Pegboard Test,⁴ where the total score was calculated as the sum of the number of pegs inserted over 30 s by (a) the left hand, (b) the right hand and (c) bimanually; (3) the TUG test,⁸ the length of time taken to arise from a chair, walk 3 metres in a straight line, turn and walk and sit back down; (4) the Flamingo Test,²⁵ where the time an individual was able to stand on their preferred leg (to a maximum of 30 s) was measured; and (5) Hoehn and Yahr (H&Y) stage assessment.²⁸ All assessments were performed on existing medication.

Variables assessed and data reporting

The date of data download was 20 June 2020. Variables assessed included the following: (1) demographic variables: age, sex and disease duration from diagnosis; (2) researcher-assessed motor variables: MDS-UPDRS III, Purdue Pegboard Test total, TUG, Flamingo and H&Y; (3) researcher-assessed non-motor variables: MDS-UPDRS part IA researcher-administered non-motor aspects of experiences of daily living and the Montreal Cognitive Assessment²⁹; and (4) participant-reported variables: MDS-UPDRS part IB self-reported non-motor aspects of experiences of daily living, MDS-UPDRS II self-reported motor aspects of experiences of daily living,¹⁵ EQ-5D-3L including the Visual Analogue Scale (VAS) rating of overall health,³⁰ Freezing of Gait Questionnaire,³¹ the number of falls in the preceding 6 months and the need for help at home.

Missing data

Complete sets of data (MDS-UPDRS III, Purdue Pegboard Test total and TUG) collected at in-person clinic visits were included in the creation of the composite clinical motor score. Incomplete sets of data were excluded from analyses (online supplemental figure 1).

Statistical analyses

All analyses were conducted using MATLAB software (VR2020a; MathWorks, USA). SPSS (V.27.0) was used to generate online supplemental figure 2.

Composite clinical motor score creation

Principal component analysis (PCA)³² was applied to complete sets of data (MDS-UPDRS III, Purdue Pegboard Test total and TUG) from the controls and individuals with iRBD and PD, with data contributed at different 18 monthly visits treated as independent, using the MATLAB function `pca`, where `[~, score, ~]=pca(input_data)`. Documentation relating to the `pca` function may be viewed at <https://uk.mathworks.com/help/stats/pca.html> and the code underlying the `pca` function may be accessed by MATLAB users by entering 'open `pca`' into the command window. With

the aim of maximising the application of our results to external data sets, an empirically chosen minimal number of three semi-quantitative variables were fed into the PCA. Given its limited use outside of the Oxford Discovery study, the Flamingo Test was excluded from the creation of the composite clinical motor score. H&Y stage was similarly excluded due to the blanket score of 0 across individuals with iRBD and controls. We used PCA as it is a commonly used unsupervised learning algorithm, whereby the first principal component captures the maximum variance in the data among all linear combinations of the constituent clinical scores. The first principal component, explaining 65.5% of the total variance, was selected to form the composite clinical motor score. Values of the first principal component were scaled through the subtraction of the minimum, division by the range and multiplication by 100, such that composite clinical motor scores lay on a scale from 0 to 100.

Evaluation schemata

Descriptive statistics, t-tests for continuous variables, χ^2 tests for categorical or ordinal variables, and Spearman’s correlation coefficient when evaluating the strength of correlations between two continuous variables were used. Where necessary, to allow for the direct comparison between the composite clinical motor score and the MDS-UPDRS III, both scores were standardised through the subtraction of their corresponding mean and division by their SD. Statistical significance was inferred from a two-sided $p < 0.05$.

An exhaustive and detailed validation was undertaken to gauge the efficacy of the proposed composite clinical motor score in providing a comprehensive overview of motor symptoms, by comparing it with the MDS-UPDRS III based on the following seven criteria: (1) *discrimination accuracy*: the ability to distinguish between the disease groups (controls, iRBD and PD), whereby the accuracy was quantified using AUC values; (2) *consistency*: using the coefficient of variation (SD/mean) as an indicator of overall variability relative to the mean^{33,34} and using the weighted mean absolute percentage change (for control participants who underwent baseline and repeat in-person assessments: defined as the sum of each participant’s absolute change in score between baseline and repeat assessments/the sum of all baseline scores, with the value expressed as a percentage), a measure akin to the weighted mean absolute percentage error (wMAPE),³⁵ which allows the quantification of forecasting errors relative to a ground truth, bypassing errors that would arise with baseline values of 0, were the mean absolute percentage change to be calculated; (3) *correlation*: measuring Spearman’s correlation with other clinical measures; (4) *sensitivity to disease stage*: change in the score values for different stages of disease severity (as assessed via the H&Y stage); (5) *longitudinal progression*: ability to track disease progression over time, for all participants with iRBD and PD—additionally, for participants with PD, the longitudinal standardised score trajectories according to baseline cluster (PD cluster 1: fast motor progression with symmetrical motor disease, poor olfaction, cognition and postural hypotension; PD cluster 2: mild motor and non-motor disease with intermediate motor progression; PD cluster 3: severe motor disease, poor psychological well-being and poor sleep with an intermediate motor progression; PD cluster 4: slow motor progression with tremor-dominant, unilateral disease), as previously described, were also compared to determine whether the trajectories were preserved across scores³⁶; (6) *relative linearity*: as calculated by the wMAPE (the sum of the absolute difference between the true and predicted scores, divided by the sum of the

true scores) of individualised linear predictions based on two or more composite clinical motor scores from the same individual at discrete time points being used to predict their composite clinical motor scores at other time points; and (7) *prediction accuracy*: ability to predict clinical outcomes, quantified using AUC values, in PD including (a) falls (at least one self-reported fall in the preceding 6 months), (b) freezing (a frequency of freezing other than ‘never’; ie, a score of at least 1 in answer to the question ‘Do you feel that your feet get glued to the floor while walking, making a turn or when trying to initiate walking (freezing)?’ on the Freezing of Gait Questionnaire), (c) cognitive impairment (a score of 1 or more on MDS-UPDRS I item 1.1), and (d) problems with self-care and performing usual activities (a score of >1 on each EQ-5D-3L item, denoting the presence of at least some problems).

Furthermore, the sample sizes necessary to detect a 50% effect in a hypothetical 18-month placebo-controlled clinical trial were calculated separately for the composite clinical motor score and the MDS-UPDRS III to allow their comparison. Calculations were based on the assumption of a 50% effect over the 18-month treatment period, a 5% (two-sided) significance level and an 80% power, and mean±SD progression as observed in Oxford Discovery cohort participants (158 PSG-diagnosed iRBD and 690 PD), where assessments in individuals with PD were performed on their typical dopaminergic medication. Sample size was allocated equally to the treatment and control groups and assumed that data from all participants were valid and usable.

RESULTS

Demographics

A total of 1497 participants contributed complete sets of data at one or more visits (number of visits: controls, 362; iRBD, 591; PD, 2898; median (IQR) years of follow-up: iRBD: 1.6 (0–3.2), PD: 3.5 (1.3–5.6)). Baseline demographics are shown in table 1. Male sex was over-represented in all three groups; the imbalance was most marked in the iRBD group, in keeping with its known epidemiology. The degree of motor impairment was greatest in PD, followed by iRBD and then the controls. At baseline, 14%

Table 1 Baseline demographics

	Control	iRBD	PD
Participants, n	316	272	909
Age	64.8 (10.1)	65.2 (8.8)	67.1 (9.6)
Male sex (%)	168 (53)	239 (88)	583 (64)
Disease duration from diagnosis	–	1.3 (1.7)	1.2 (0.9)
MDS-UPDRS III*	1.75 (2.7)	4.7 (4.1)	26.3 (10.7)
Purdue Pegboard Test total†	37.2 (6.8)	35.3 (7.3)	28.7 (6.8)
Timed Up and Go‡	8.1 (1.7)	8.5 (2.3)	9.7 (3.4)
Hoehn and Yahr§	0.0 (0.0)	0.0 (0.0)	1.8 (0.5)
Montreal Cognitive Assessment¶	26.7 (2.5)	25.1 (2.8)	24.9 (3.3)

Clinical and demographic data are presented as mean and SD (in brackets). 98% of the participants identified their ethnicity as white.

*MDS-UPDRS III missing in 4 participants with iRBD and 11 participants with PD.

†6 PD missing Purdue Pegboard Test total.

‡23 participants with PD missing Timed Up and Go.

§Hoehn and Yahr stage missing in 2 controls and 3 participants with PD.

¶Montreal Cognitive Assessment missing in 5 controls, 2 participants with iRBD and 13 participants with PD.

iRBD, isolated rapid eye movement sleep behaviour disorder; MDS-UPDRS III, Movement Disorders Society-Unified Parkinson’s Disease Rating Scale III; PD, Parkinson’s disease.

Table 2 Comparative accuracies in group separation at baseline

	iRBD vs controls	PD vs controls	PD vs iRBD
Composite clinical motor score	0.63 (0.58 to 0.67)	0.94 (0.92 to 0.95)	0.89 (0.86 to 0.91)
MDS-UPDRS III	0.76 (0.72 to 0.80)	1.0 (0.99 to 1.00)	0.98 (0.98 to 0.99)
Purdue Pegboard Test	0.58 (0.53 to 0.63)	0.81 (0.79 to 0.84)	0.75 (0.72 to 0.78)
Timed Up and Go	0.54 (0.49 to 0.58)	0.68 (0.64 to 0.71)	0.64 (0.60 to 0.68)

Values indicate the area under the curve value and 95% CI as derived through a bootstrapping approach.

At baseline, the ranges (mean (SD)) in composite clinical motor score were as follows: controls: 1.4–31.4 (13.2 (4.2)); iRBD: 0–30.4 (16.5 (5.5)); PD: 7.5–65.5 (26.0 (7.7)).

iRBD, isolated rapid eye movement sleep behaviour disorder; MDS-UPDRS III, Movement Disorders Society-Unified Parkinson’s Disease Rating Scale III; PD, Parkinson’s disease.

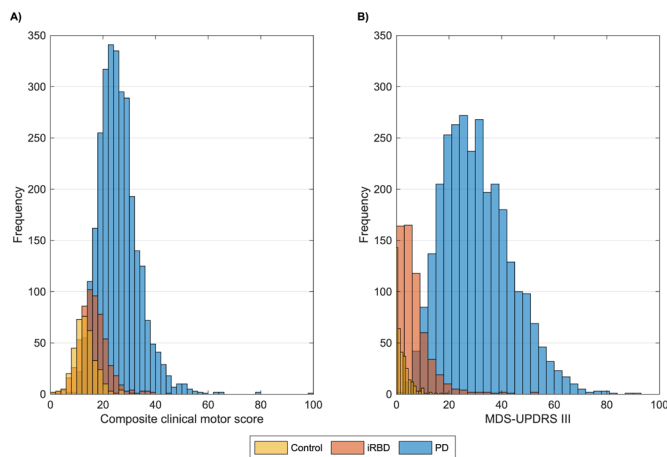


Figure 1 Histograms illustrating the distribution by disease group of (A) composite clinical motor scores and (B) MDS-UPDRS III scores. iRBD, isolated rapid eye movement sleep behaviour disorder; MDS-UPDRS III, Movement Disorders Society-Unified Parkinson’s Disease Rating Scale III; PD, Parkinson’s disease.

of participants with PD were drug-naïve; by the fifth 18-month visit, all but one (who wished to remain off medication as long as possible) were on dopaminergic medication. The mean (SD) change in levodopa equivalent daily dose³⁷ per visit was 124 (131). Of the 899 participants with PD who had undergone clinical baseline stratification as previously described,³⁶ 295 (33%) were in cluster 1, 153 (17%) in cluster 2, 214 (24%) in cluster 3, and 237 (26%) in cluster 4.

Composite clinical motor score calculation

The following formula allows the calculation of the composite clinical motor score from constituent MDS-UPDRS III, Purdue Pegboard Test and TUG scores:

$$\text{Composite clinical motor score} = ((\text{MDS-UPDRS III} - 2.3.8829) \times 0.0377 + (-1 \times \text{Purdue} + 28.9286) \times 0.0739 + (\text{TUG} - 9.8746) \times 0.1414 + 3.8239) \times 6.0946.$$

A spreadsheet calculator is provided in online supplemental material to ease calculation.

Group separation at baseline

Through the identification and application of simple thresholds to distinguish groups, the composite clinical motor score was associated with greater accuracies in discriminating iRBD or PD from the controls and PD from iRBD than either the Purdue Pegboard Test or TUG, although accuracies were less than for the MDS-UPDRS III (table 2 and online supplemental table 1). Compared with the MDS-UPDRS III where a

floor effect was observed, the composite clinical motor score showed a wider score distribution for iRBD and controls (figure 1). No ceiling effect was observed for both scores.

Consistency

The coefficient of variation, across data from all participants, associated with the composite clinical motor score (37%) was lower than that associated with MDS-UPDRS parts II (participant-reported motor severity) (78%) and III (subjective clinician-rated motor severity) (67%), indicating lower variability relative to the mean; values were similar to those associated with semiquantitative measures of motor severity, namely the Purdue Pegboard Test (28%) and TUG test (38%).

Forty-six control participants returned for a second in-person assessment and contributed data allowing the calculation of a composite clinical motor score. The second in-person control assessment was performed a mean (SD) of 3.7 (1.4) years after the baseline visit. From baseline, there was a mean (SD) increase in MDS-UPDRS III score of 0.7 (1.3) points per year compared with baseline scores of 2.2 (2.5). Comparatively, the mean (SD) increase in composite clinical motor score was 0.2 (0.9) points per year compared with baseline scores of 13.7 (4.3). The weighted mean absolute percentage change in the MDS-UPDRS III was 142% compared with 19% in the composite clinical motor score.

Correlation with other clinical measures

Significant (p<0.001) correlations were evident between the MDS-UPDRS III and the Purdue Pegboard Test (r=−0.58) and the MDS-UPDRS III and the TUG (r=0.44). The correlation coefficients (p<0.001) between the Purdue Pegboard Test and MDS-UPDRS III finger tapping and hand movement subitems ranged from −0.44 to −0.41; those between the TUG and MDS-UPDRS III arising from chair and gait subitems were 0.54 and 0.48, respectively.

Unsurprisingly, high correlation coefficients (absolute r=0.70–0.86) were observed between the composite clinical motor score and its constituent clinical scores (MDS-UPDRS III, Purdue Pegboard Test and TUG). Correlations with the Flamingo Test, MDS-UPDRS II and EQ-5D-3L VAS rating of overall health were higher with the composite clinical motor score compared with the MDS-UPDRS III (figure 2).

Change with PD disease severity

Across disease groups, compared with the standardised MDS-UPDRS III (scaled by subtracting the mean then dividing by the SD to facilitate the direct comparison between scores), the IQR of standardised composite clinical motor scores at baseline was larger (wider boxes) for lower H&Y stages (online supplemental figure 2), detecting greater variation, particularly in milder disease.

Ability to track disease progression on a group level

A wider score range for a given disease duration was evident particularly in early iRBD, where additionally the magnitude of change over time was greater with the composite clinical motor score compared with the MDS-UPDRS III (figure 3A). There was no difference between standardised composite clinical motor score and MDS-UPDRS III score ranges for a given PD disease duration (figure 3B) or in linearly fitted longitudinal standardised score trajectories according to baseline cluster (online supplemental figure 3 and online supplemental table 2).

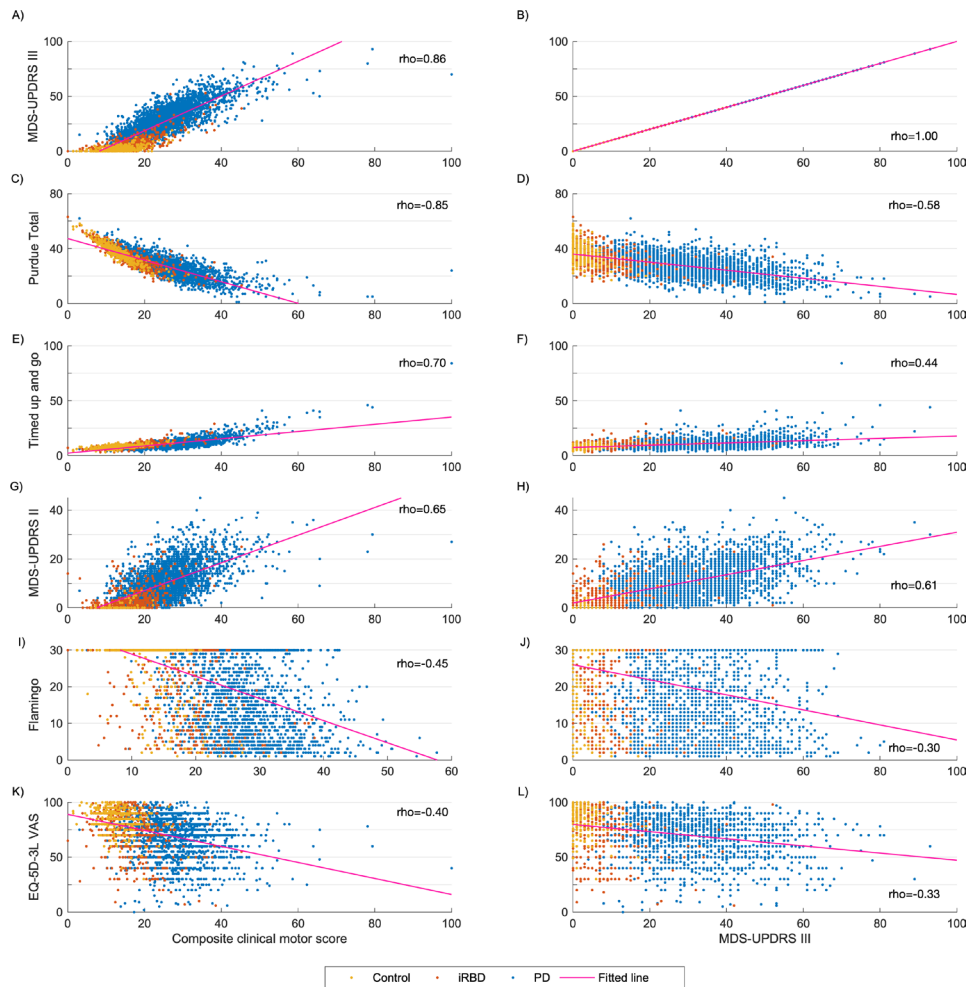


Figure 2 Scatter plots comparing the relationship between the composite clinical motor scores (A, C, E, G, I, K) and MDS-UPDRS III (B, D, F, H, J, L) scores with other clinical scores. iRBD, isolated rapid eye movement sleep behaviour disorder; MDS-UPDRS III, Movement Disorders Society-Unified Parkinson’s Disease Rating Scale III; VAS, Visual Analogue Scale; PD, Parkinson’s disease.

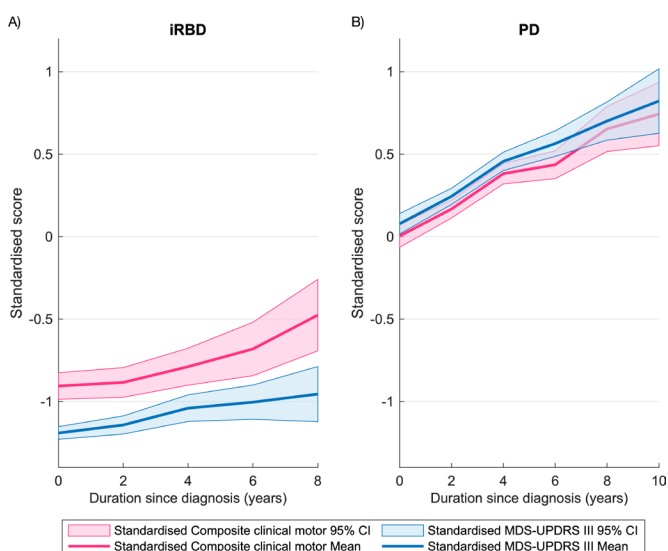


Figure 3 Longitudinal change in standardised composite clinical motor and MDS-UPDRS III scores in individuals with a) iRBD and b) PD. iRBD, isolated rapid eye movement sleep behaviour disorder; MDS-UPDRS III, Movement Disorders Society-Unified Parkinson’s Disease Rating Scale III; PD, Parkinson’s disease.

Relative score linearity on an individual level

Ninety-three individuals with iRBD and 564 individuals with PD contributed data at three or more visits. The wMAPE of individualised linear predictions for the composite clinical motor score was around half that of the MDS-UPDRS III (composite wMAPE: iRBD 22%, PD 18%; MDS-UPDRS III wMAPE: iRBD 60%, PD 34%) (online supplemental figure 4), suggesting greater score linearity.

Clinical outcome prediction

Falls, freezing, problems with self-care and usual activities were predicted with significantly greater accuracy by the composite clinical motor score compared with the MDS-UPDRS III score (table 3).

	Composite clinical motor score	MDS-UPDRS III
Falls*	0.70 (0.67 to 0.73)	0.64 (0.61 to 0.67)
Freezing*	0.72 (0.69 to 0.75)	0.66 (0.63 to 0.69)
Cognitive impairment	0.63 (0.61 to 0.65)	0.59 (0.57 to 0.61)
Self-care*	0.75 (0.73 to 0.77)	0.70 (0.68 to 0.72)
Usual activities*	0.69 (0.67 to 0.71)	0.64 (0.62 to 0.66)
*Non-overlapping 95% CI.		
MDS-UPDRS III, Movement Disorders Society-Unified Parkinson’s Disease Rating Scale III.		

Table 4 Sample size estimations for detection of a 50% effect in a hypothetical 18-month placebo-controlled clinical trial

Population	Primary endpoint	18-month change from baseline (CV, %)	Total sample size (per group)	% benefit of using the composite clinical motor score
PD	MDS-UPDRS part III	2.96±10.04 (339.19)	1450 (725)	64% fewer participants
	Composite clinical motor score	2.37±4.85 (204.64)	528 (264)	
iRBD	MDS-UPDRS part III	3.08±7.74 (251.30)	794 (397)	51% fewer participants
	Composite clinical motor score	2.52±4.43 (175.79)	392 (196)	

CV, coefficient of variation in percentage defined as the ratio of SD to the mean (average); iRBD, isolated rapid eye movement sleep behaviour disorder; MDS-UPDRS III, Movement Disorders Society-Unified Parkinson's Disease Rating Scale III; PD, Parkinson's disease.

Sample size calculation

For detection of a 50% effect in a hypothetical 18-month placebo-controlled clinical trial, 64% fewer participants with PD and 51% fewer participants with iRBD were estimated to be required using the composite clinical motor score as an outcome measure compared with the MDS-UPDRS III (table 4).

DISCUSSION

There remains a critical need for novel clinical trial outcome measures, with the sensitivity to detect subtle change, improving on the subjectivity of existing, yet popular, clinical scales. We describe the creation of a composite clinical motor score, rooted in detailed clinical assessments from a combined total of 1497 individuals, including those with iRBD, PD and controls. Compared with the MDS-UPDRS III, the composite clinical motor score is associated with (1) slightly worse group discrimination accuracies; (2) lower variability relative to the mean and a lower percentage change relative to baseline on repeat assessment in control participants, indicating improved score consistency; (3) equivalent or superior correlation with other clinical measures; (4) greater score variability for a given PD stage (particularly in milder disease); (5) greater score variability for a given disease duration in iRBD and magnitude of change over time; (6) superior accuracy of individualised longitudinal linear score predictions; and (7) superior accuracy in the prediction of clinical outcomes of falls, freezing, problems with self-care and usual activities. Sixty-four per cent fewer participants with PD and 51% fewer participants with iRBD were estimated to be required for clinical trials using the composite clinical motor score as an outcome measure compared with the MDS-UPDRS III.

A greater sensitivity of the composite clinical motor score to detect subtle motor change is suggested by its greater magnitude of change over time as well as its greater score variability in iRBD and in the milder stages of PD, relative to the MDS-UPDRS III. Acknowledging that true disease progression is not necessarily linear,^{38 39} although an assumption made by many studies,^{36 40–42} the superior accuracy of linear predictions applied to the composite clinical motor score suggests greater consistency of measurement, with greater within-individual MDS-UPDRS III score variability over time being attributed to measurement inaccuracies rather than true disease variation alone. A reflection of the consistency of the composite clinical motor score, significant reductions were seen in the sample sizes necessary to demonstrate treatment effects compared with MDS-UPDRS III, suggesting promise for future clinical trial use. This is particularly pertinent for emerging adaptive trial designs (eg, multiarm, multistage), whereby interim analyses at short time intervals resulting in different arm modification or closure due to lack of efficacy are needed.⁴³

Limitations

While created from one of the largest deeply phenotyped control, iRBD and PD data sets worldwide, the composite clinical motor score will benefit from further evaluation in external data sets. Although our results so far suggest a particular strength of the composite clinical motor score in the quantification of mild motor impairment, its relevance to individuals with advanced disease remains to be elucidated. As with each of its constituent scores (MDS-UPDRS III, Purdue Pegboard Test and TUG), the composite clinical motor score quantifies motor impairment independent of its aetiology and any effect of ageing. With the greater ability of the composite clinical motor score to detect subtle motor change, motor impairment is detected in individuals who would otherwise have been awarded a score of 0 using the MDS-UPDRS III; the trade-off is a reduction in group separation accuracies due to the increase in overlap between the different groups. In PD, score change over time is evaluated using longitudinal data relating to a median follow-up of 3.5 (Q1–Q3: 1.3–5.6) years. However, the absence of paired off/on assessments within the Discovery study, across *all* clinical assessments of motor function, prevented the assessment of the natural history of motor impairment in PD and medication response using the composite clinical motor score; determining the score difference associated with clinically meaningful change will be important if it is to be considered as an outcome measure in future therapeutic trials. While the superior ability of the composite clinical motor score in predicting clinical outcomes in PD adds to its value, it remains to be seen whether important clinical milestones, such as the point and nature of phenoconversion, may also be predicted in individuals with iRBD.

Barriers to the application of the composite clinical motor score in routine clinical practice include the necessity to measure all three constituents (MDS-UPDRS III, Purdue Pegboard Test and TUG); of questionable feasibility considering typical time constraints, though used by a number of research studies to date.^{22 44} Refinement of the composite clinical motor score, through the identification and exclusion of redundant constituents or their subitems, may be the subject of future work. One approach we are investigating to directly address the time constraint concern is the use of smartphone motor testing⁴⁵ (involving the automated 10 min assessment of voice, balance, gait, reaction time, dexterity, rest tremor and postural tremor) to predict the composite clinical motor score. In doing so, it is hoped that clinicians and patients alike may be equipped with the tools to derive objective measures of motor severity, not only in clinics but at home, paving the way to more informed personalised treatment decisions and potentially increasing the engagement of patients in their care.

CONCLUSION

The composite clinical motor score combines the MDS-UPDRS III, Purdue Pegboard Test and TUG, three clinical assessments of

motor severity, to yield a score that demonstrates greater consistency and sensitivity to change than the existing gold standard outcome measure. Further work is indicated to explore its application to clinical trials.

Acknowledgements All of this work would not have been possible were it not for the time and effort invested by participants in the Oxford Discovery research study along with members of the research team at the different sites. Additionally, we are grateful to Jonas Wiedemann, a biostatistician at Lundbeck, for his help with the sample size calculations.

Contributors All authors have been involved in revising this manuscript critically for intellectual content and have read and approved the final version of the manuscript. The paper has not been previously published and is not under simultaneous consideration by another journal. MH accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding This study was funded by the Monument Trust Discovery Award from Parkinson's UK (J-1403) and supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC).

Disclaimer The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The study sponsor had no role in the study design; in the collection, analysis and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication.

Competing interests CL was funded by the NIHR Oxford BRC. TB was funded by the Wellcome Trust Doctoral Training Fellowship. TQ is currently a short term Consultant for Jazz Pharmaceuticals and a CI for an NIHR grant for a therapeutic trial in OSA.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the South Central - Oxford A Research Ethics Committee (REC references: 16/SC/0108 and 10/H0505/71). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Qualified investigators seeking access to de-identified participant data relating to the Oxford Discovery cohort may submit their request by means of a formal application to the OPDC Data Access Committee. The application form, protocol, and terms and conditions may be found at opdc.medsci.ox.ac.uk/external-collaborations.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Christine Lo <http://orcid.org/0000-0002-5671-3881>

Michael Lawton <http://orcid.org/0000-0002-3419-0354>

Michele Tao-Ming Hu <http://orcid.org/0000-0001-6382-5841>

REFERENCES

- Athauda D, Foltynie T. The ongoing pursuit of neuroprotective therapies in Parkinson disease. *Nat Rev Neurol* 2015;11:25–40.
- Athauda D, Maclagan K, Skene SS, et al. Exenatide once Weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *Lancet* 2017;390:1664–75.
- Initiative CTT. Use case for developing novel endpoints generated using mobile technology: Parkinson's disease 2017:1–7.
- Tiffin J, Asher EJ. The Purdue pegboard; norms and studies of reliability and validity. *J Appl Psychol* 1948;32:234–47.
- Proud EL, Bilney B, Miller KJ, et al. Measuring hand dexterity in people with Parkinson's disease: reliability of Pegboard tests. *Am J Occup Ther* 2019;73:7304205050p1–7304205050.
- Růžicka E, Krupička R, Zárubová K, et al. Tests of manual dexterity and speed in Parkinson's disease: not all measure the same. *Parkinsonism Relat Disord* 2016;28:118–23.
- Darweesh SKL, Wolters FJ, Hofman A, et al. Simple test of manual dexterity can help to identify persons at high risk for neurodegenerative diseases in the community. *J Gerontol A Biol Sci Med Sci* 2017;72:75–81.
- Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991;39:142–8.
- Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease. *Phys Ther* 2001;81:810–8.
- van Lummel RC, Walgaard S, Hobert MA, et al. Intra-Rater, inter-rater and test-retest reliability of an instrumented timed up and go (iTUG) test in patients with Parkinson's disease. *PLoS One* 2016;11:e0151881.
- Haas B, Clarke E, Elver L, et al. The reliability and validity of the L-test in people with Parkinson's disease. *Physiotherapy* 2019;105:84–9.
- Ferrazzoli D, Ortelli P, Riboldazzi G, et al. Effectiveness of Rotigotine plus intensive and goal-based rehabilitation versus Rotigotine alone in "de-novo" Parkinsonian subjects: a randomized controlled trial with 18-month follow-up. *J Neurol* 2018;265:906–16.
- Zhou H, Wang L, Zhang C, et al. Acute effects of subthalamic deep brain stimulation on motor outcomes in Parkinson's disease; 13 year follow up. *Front Neurol* 2019;10:689.
- Fahn S, Elton RL. UPDRS program members. Unified Parkinsons Disease Rating Scale. In: Fahn S, Marsden CD, Goldstein M, et al, eds. *Recent developments in Parkinsons disease*. 2. Florham Park, NJ: Macmillan Healthcare Information, 1987.
- 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;23:2129–70.
- Siderowf A, McDermott M, Kiebertz 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;17:758–63.
- Post B, Merkus MP, de Bie RMA, et al. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Mov Disord* 2005;20:1577–84.
- Richards M, Marder K, Cote L, et al. Interrater reliability of the unified Parkinson's disease rating scale motor examination. *Mov Disord* 1994;9:89–91.
- Horváth K, Aschermann Z, Ács P, et al. Minimal clinically important difference on the motor examination part of MDS-UPDRS. *Parkinsonism Relat Disord* 2015;21:1421–6.
- Evers LJW, Krijthe JH, Meinders MJ, et al. Measuring Parkinson's disease over time: the real-world within-subject reliability of the MDS-UPDRS. *Mov Disord* 2019;34:1480–7.
- Regnault A, Borojerdi B, Meunier J, et al. Does the MDS-UPDRS provide the precision to assess progression in early Parkinson's disease? Learnings from the Parkinson's progression marker initiative cohort. *J Neurol* 2019;266:1927–36.
- Postuma RB, Iranzo A, Hu M, et al. Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain* 2019;142:744–59.
- Silveira CRA, MacKinley J, Coleman K, et al. Amroxol as a novel disease-modifying treatment for Parkinson's disease dementia: protocol for a single-centre, randomized, double-blind, placebo-controlled trial. *BMC Neurol* 2019;19:20.
- Szewczyk-Krolikowski K, Tomlinson P, Nithi K, et al. The influence of age and gender on motor and non-motor features of early Parkinson's disease: initial findings from the Oxford Parkinson disease center (Opdc) discovery cohort. *Parkinsonism Relat Disord* 2014;20:99–105.
- Barber TR, Lawton M, Rolinski M, et al. Prodromal parkinsonism and neurodegenerative risk stratification in REM sleep behavior disorder. *Sleep* 2017;40:zsx071.
- International classification of sleep disorders third edition. *American Academy of Sleep Medicine* 2014.
- Hughes AJ, Daniel SE, Kilford L, et al. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992;55:181–4.
- Hoeft MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967;17:427–42.
- Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695–9.
- EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- Giladi N, Shabtai H, Simon ES, et al. Construction of freezing of gait questionnaire for patients with parkinsonism. *Parkinsonism Relat Disord* 2000;6:165–70.
- Jolliffe IT. *Principal component analysis*. 2 ed. New York, NY: Springer, 2002.
- Usui T, Macleod MR, McCann SK, et al. Meta-Analysis of variation suggests that embracing variability improves both replicability and generalizability in preclinical research. *PLoS Biol* 2021;19:e3001009
- Kiyomi F, Nishikawa M, Yoshida Y, et al. Comparison of intra-individual coefficients of variation on the paired sampling data when inter-individual variations are different between measures. *BMC Res Notes* 2016;9:115.

- 35 Horng S, O'Donoghue A, Dechen T, *et al.* Secondary use of COVID-19 symptom incidence among hospital employees as an example of syndromic surveillance of hospital admissions within 7 days. *JAMA Netw Open* 2021;4:e2113782
- 36 Lawton M, Ben-Shlomo Y, May MT, *et al.* Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J Neurol Neurosurg Psychiatry* 2018;89:1279–87.
- 37 Tomlinson CL, Stowe R, Patel S, *et al.* Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov Disord* 2010;25:2649–53.
- 38 Reinoso G, Allen JC, Au W-L, *et al.* Clinical evolution of Parkinson's disease and prognostic factors affecting motor progression: 9-year follow-up study. *Eur J Neurol* 2015;22:457–63.
- 39 Prange S, Danaila T, Laurencin C, *et al.* Age and time course of long-term motor and nonmotor complications in Parkinson disease. *Neurology* 2019;92:e148–60.
- 40 Holden SK, Finseth T, Sillau SH, *et al.* Progression of MDS-UPDRS scores over five years in de novo Parkinson disease from the Parkinson's progression markers initiative cohort. *Mov Disord Clin Pract* 2018;5:47–53.
- 41 Vu TC, Nutt JG, Holford NHG. Progression of motor and nonmotor features of Parkinson's disease and their response to treatment. *Br J Clin Pharmacol* 2012;74:267–83.
- 42 Postuma RB, Lang AE, Gagnon JF, *et al.* How does parkinsonism start? prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain* 2012;135:1860–70.
- 43 Barthel FM-S, Parmar MKB, Royston P. How do multi-stage, multi-arm trials compare to the traditional two-arm parallel group design--a reanalysis of 4 trials. *Trials* 2009;10:21.
- 44 Hartelt E, Scherbaum R, Kinkel M, *et al.* Parkinson's disease multimodal complex treatment (PD-MCT): analysis of therapeutic effects and predictors for improvement. *J Clin Med* 2020;9. doi:10.3390/jcm9061874. [Epub ahead of print: 16 06 2020].
- 45 Arora S, Baig F, Lo C, *et al.* Smartphone motor testing to distinguish idiopathic REM sleep behavior disorder, controls, and PD. *Neurology* 2018;91:e1528–38.