RESEARCH ARTICLE

Item Response Theory Analysis of the MDS-UPDRS Motor Examination: Tremor vs. Nontremor Items

Michelle Hyczy de Siqueira Tosin, RN, MSN,[1,2] Christopher G. Goetz, MD,[2] Sheng Luo, PhD,[3] Dongrak Choi, MS,[3] and Glenn T. Stebbins, PhD[2*]

[1]*Department of Nursing, Federal Fluminense University, Niterói, Rio de Janeiro, Brazil*
[2]*Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA*
[3]*Department of Biostatistics and Bioinformatics, Duke University, Medical Center, Durham, North Carolina, USA*

ABSTRACT: Background: In PD, tremor severity behaves differently from other core motor features. However, the most commonly used assessment of overall motor severity, total MDS-UPDRS Motor Examination (Part 3) score, does not account for this distinction.
Objectives: To investigate the Motor Examination (Part 3) using Item Response Theory approaches focusing on sample-independent strategies that assess how well items measure latent models of PD motor severity.
Methods: Data from 6,298 PD patients were analyzed with graded response model Item Response Theory approaches involving two analyses all 33 Part 3 items versus the 10 tremor items and 23 bradykinesia, rigidity, gait, and posture items considered separately. The strength of relationship between items and the latent measure of parkinsonian motor severity (discrimination parameter) and calculated thresholds (location parameters) were assessed using the *mirt* program implemented in R (R Foundation for Statistical Computing, Vienna, Austria).
Results: Analyzing all Part 3 items together, nontremor items demonstrated good discrimination parameters (mean = 1.83 ± 0.37) and range of thresholds (–1.73 to +4.42), but tremor items had poor discrimination (mean = 0.52 ± 0.76) and thresholds (–0.69 to 14.29). Segregating nontremor from tremor items in two independent analyses provided markedly improved discrimination and location parameters for both.
Conclusions: MDS-UPDRS Part 3 tremor and nontremor items have very different relations to the construct of PD severity. Strongly improved clinimetric properties for Part 3 are obtained when tremor and nontremor items are considered separately. We suggest that evaluating PD motor severity, as an operationalized summary measure, is best attained through separate analyses with tremor and nontremor motor scores. © 2020 International Parkinson and Movement Disorder Society

Key Words: clinimetrics; item response theory; MDS-UPDRS; Parkinson's disease

Most rating scales are designed to consider the key components of a given disease and provide an operational ranking of the intensity of impairment or disability related to each component.[1] Always residing behind these measures are the conceptual framework of the disease itself and its severity, so that a summary severity score of the components provides an overall index of the disease severity. This method is particularly logical and effective when all components of the disease relate to a common pathophysiology. In Parkinson's disease (PD), there is a long history of rating scale application, but also a historical acknowledgment that tremor severity is often quite different than other core features of the disease, namely bradykinesia, rigidity, posture, gait, and balance impairment severities. In clinical presentation, disease progression, pharmacological response, surgical response, and relation with other nonmotor attributes of PD, tremor severity

*Correspondence to: Dr. Glenn T. Stebbins, Department of Neurological Sciences, Section of Movement Disorders, Rush University Medical Center, 1725 West Harrison Street, Suite 755, Chicago, IL 60612, USA; E-mail: gstebbin@rush.edu

Funding agencies: The Rush Parkinson's Disease and Movement Disorders Program is a designated Clinical Center of Excellence supported by the Parkinson Foundation.

Relevant conflicts of interest/financial disclosures: Nothing to report.

Full financial disclosures and author roles may be found in the online version of this article.

Received: 3 February 2020; **Revised:** 23 April 2020; **Accepted:** 28 April 2020

Published online 00 Month 2020 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.28110

has continually been considered as potentially distinct from nontremor PD manifestations.[2]

In this context, a core rating scale issue for PD that needs refinement is how to measure the construct of overall disease severity when the disease itself has the potential to have two archetypically different phenotypic components, each potentially with its own severity. Most pronounced in the two polarized designations "tremor-predominant" and "akinetic-rigid" PD,[2] the problem of arriving at a total motor assessment of PD severity in the context of dichotomies between tremor and nontremor rating items exists for rating all patients with PD.

The International Parkinson and Movement Disorder Society revision of the UPDRS (MDS-UPDRS) has become the most widely used scale for measuring parkinsonian symptoms in clinical and research practice.[3] The scale has extensive clinimetric support, but the original validation studies of the MDS-UPDRS, and most of the subsequent examinations, were conducted in the framework of Classic Test Theory (CTT).[4-8] This framework examines item-level relationships and the way items group together to assess domains.[9] With CTT, clinimetric properties of reliability are commonly examined by assessing the strength of the relationship of individual items to the total scale (item-to-total correlations), amount of scale variance attributable to the influence of individual items (Cronbach's alpha), and reliability of item scores across different observers (inter-rater reliability) or across observations (intrarater reliability). Validity assessments typically examine the domain coverage of the items (content validity), dimensionality of the items (construct validity), or the relationship of scores to other measures of similar constructs (concurrent, convergent, or criterion validity).[10]

Although CTT-based studies of the MDS-UPDRS have been fruitful, the framework is limited by a reliance on the relationship of the items within a given sample-dependent context. Because of this dependency, varying results will likely emerge when different samples are examined. An alternate, but more complex, approach to examining clinimetric properties is Item Response Theory (IRT).[11]

The framework for IRT differs from CTT in that the measurement characteristics of the items and scores are calibrated to a sample-independent latent model of disease severity.[12] Whereas CTT approaches may demonstrate items clustering into a single factor relating to the total score, the IRT approach examines the relationship of item scoring to a latent model of disease severity. This approach allows for person-invariant examination of item- and scale-score relations to the magnitude of the latent model of severity, by differentiating individual phenotypic characteristics from symptoms related to an individual's underlying latent (or hidden) variable to the pattern of responses to the items on the scale.[13] Importantly, if items or item clusters chart disease severity differently than other items or clusters, IRT will identify these patterns, allowing clinicians to consider strategies of data analysis to rectify the statistical discord, leveraging the metrics provided by the clinical rating scales.[14]

The use of IRT as an approach to examine clinimetric properties of the MDS-UPDRS is recent, with methodological variability in the IRT models used, characteristics of PD samples studied, and portions of the scale used in the analyses. Most of these studies use longitudinal data from the Parkinson's Progression Marker initiative,[15-18] a cohort of early-stage PD patients who do not require treatment at baseline and are followed longitudinally. The majority of these studies applied IRT methods to examine longitudinal change of either total scale scores or unique combinations of different items of the MDS-UPDRS. None of these studies examined a cross-sectional cohort of PD patients with the full gamut of disease severity, using traditional scoring methods for the most commonly used outcome measure,[3] the Motor Examination of the MDS-UPDRS.

The purpose of the present study was to apply a graded response IRT model to analyze the items of the MDS-UPDRS Part 3 Motor Examination scores as typically applied in clinical and research settings to a large, cross-sectional, international sample of patients with PD. This allowed us to test whether overall motor severity is best captured with a single score or with tremor and nontremor severities calculated separately. In the latter case, considering the tremor and nontremor scores separately might provide more sensitive measures depending on the context.

Patients and Methods

Sample

The data from the MDS-UPDRS translation program were used in this study to analyze the 33 items of the MDS-UPDRS Part 3 Motor Examination. The MDS-UPDRS translation program,[19] initiated in 2007, is an ongoing international program sponsored by the MDS designed to develop and validate translations of the MDS-UPDRS. To date, the scale has been successfully translated into 14 languages. The process involves translation into the native language; independent back-translation from the native language into English; cultural harmonization through cognitive pretesting; and validation of the translation requiring a sample of at least 350 complete examinations using the translated scale. Examiners for the translated scale are required to complete the MDS-UPDRS online training certificate program offer by the MDS (https://www.movementdisorders.org/MDS/MDS-Rating-Scales/Training-Programs.htm). Those translations meeting established criteria for

confirmatory factor analyses are considered validated MDS-UPDRS versions.[19] The data are collected at only one time point, and no longitudinal follow-up is conducted.

This study consisted of 6,684 records from a cohort that included a heterogeneous sample of patients diagnosed with PD and considered native speakers from 14 different languages. The inclusion criteria included a diagnosis of PD according to UK Brain Bank criteria and fluent speakers of the native language. Participants were excluded if they were not diagnosed with PD or were not native speakers of the language of the translation. The MDS-UPDRS was administered in a single visit. Demographic data (age, sex, and race) and disease history (duration of PD) data were collected for all subjects. Information on the use of medications for treating the symptoms of PD was not available.

Statistical Analysis

A graded-response IRT approach was conducted using the multidimensional IRT *mirt* R statistical program (R Foundation for Statistical Computing, Vienna, Austria).[20] In the context of Part 3 of the MDS-UPDRS, IRT approaches use a categorical model to develop a latent-trait measure of overall parkinsonian motor severity, representing the underlying disease severity. Applying a graded-response model, a particular response option of each item within the scale can be assessed for its relationship to this latent model and the probability of response for a person with trait level. The relationship was determined by five parameters—the discrimination parameter and the four location parameters. The discrimination parameter corresponds to the inverse of the residual variability of the item, where higher discrimination value means that the item is more powerful for determining the individual's overall parkinsonian motor severity. The magnitude of the discrimination parameter import can be judged using the following criteria: none = 0; very low = 0.01 to 0.34; low = 0.35 to 0.64; moderate = 0.65 to 1.34; high = 1.35 to 1.69; very high = >1.70; and perfect = $+\infty$.[21] The location parameters—also called difficulty parameters—indicate the probability threshold for the transitioning of an item score from 0 to 1 (from normal to slight), from 1 to 2 (from slight to mild), from 2 to 3 (from mild to moderate), and from 3 to 4 (from moderate to severe) in relation to the latent-trait measure of overall parkinsonian motor severity. The relationship between item score transition probability and the latent measure of overall parkinsonian motor severity can be graphically represented in item characteristic curves (ICCs). In this study, this method was used to measure changes in the individual item scaling profiles, and it was expected that the location for each ascending transition will increase along the latent measure of parkinsonian motor severity (see the Supporting Information for equations utilized in the *mirt* program for calculating discrimination and location parameters).

We conducted two separate IRT analyses. The first analysis explored the discrimination and location parameters for all 33 items from the MDS-UPDRS Part 3 when considered as a unidimensional model. The second analysis looked at the 10 items measuring tremor and the 23 items measuring bradykinesia, rigidity, gait, and posture, when considered as a multidimensional model. We compared the unidimensional to the multidimensional model using the likelihood ratio test (χ^2). The Bayesian information criterion (BIC) was used as transformations of the likelihood ratio test, where lower values indicate better fit to the model[22] (see the Supporting Information for examples of the *mirt* programming for these analyses).

To assess the ability to estimate the scores of the 10 tremor items from both the unidimensional model and the multidimensional model, we adopted various graphical diagnostics. Specifically, the mirror plots compare the observed proportion of response in each category of each tremor item (one plot for each item) to the proportion estimated from models. The scatter plots compare these proportions of all 10 tremor items. To assess the goodness of fit, we compare the fit obtained based on the predicted ICCs for each tremor item from the unidimensional model to the fit obtained from the multidimensional model and displayed these effects in goodness-of-fit plots.

Results

From an original set of 6,684 records of MDS-UPDRS Part 3 scores, our study utilized the 6,382 records with full data on all 33 items. The study sample included 53% men, a mean age of 65.6 ($\pm$10.6), mostly white (74%), and an average education of 11 years ($\pm$4.8). The mean duration of PD was 7.7 ($\pm$5.9) years. The mean MDS-UPDRS Part 3 score was 33.7 ($\pm$18.3) with the prevalence of individuals in the early and moderate stages of the disease (H & Y Stage Median: 2; range, 1–5; Table 1).

We first analyzed all 33 items of Part 3 of the MDS-UPDSR in a unidimensional model. This analysis required 54 iterations (goodness-of-fit: log-likelihood = –215,933; BIC = 433,312), and the discrimination and threshold values for all 33 items are presented in Table 2. The discrimination parameters were "very high" for items measuring bradykinesia, arising from chair, gait, and global spontaneity of movement, and "high" for items measuring rigidity and posture (mean, 1.83 $\pm$ 0.37). Items measuring tremor had "very low" and "low" discrimination scores (mean, 0.52 $\pm$ 0.08). The location parameters for

TABLE 1. Demographic characteristics of the sample

Variable	PD Cohort (n = 6,382)
Sex, N (%)	
Male	53
Female	47
Age, years	
Mean (SD)	65.6 (±10.6)
Race, N (%)	
White	74
Education, years	
Mean (SD)	11.1 (±4.8)
PD duration, years	
Mean (SD)	7.7 (±5.9)
MDS-UPDRS-Part 3	33.7 (±18.3)
Mean (SD)	
H & Y stage, N (%)	
1	13
2	49
3	27
4	9
5	2

SD, standard deviation.

the 33 items reflected a sizeable range of underlying symptom severity with a range for nontremor items from –1.73 to 4.42 and a range of –0.69 to 14.29 for tremor items. Threshold values are represented in the ICCs (see Supporting Information Fig. S1A).

Given the marked difference in discrimination and location parameters between the nontremor and tremor items, we conducted two additional analyses. The first was to assess whether a multidimensional model using two dimensions resulted in an improved fit. The second was to conduct a factor analysis of the two dimensions using an unweighted least squares extraction with a varimax rotation. The multidimensional analysis required 250 iterations (goodness-of-fit: log-likelihood = –207462; BIC = 416,650). Comparison of the unidimensional versus the multidimensional models indicated a significantly superior fit for the multidimensional model (χ^2 = 16,942.75; df = 32; $P \leq 0.0005$). Exploration of the structure of the multidimensional model revealed two parsimonious factors,

TABLE 2. Discrimination (Discrim) and item location parameters for all MDS-UPDRS Part 3 Items in the unidimensional model

Items		Discrim	Item Location					
			0 to 1	1 to 2	2 to 3	3 to 4	Mean	SD
3.1	Speech	1.39	–0.97	0.91	2.41	4.08	1.61	2.15
3.2	Facial expression	1.43	–1.73	0.28	2.10	3.81	1.12	2.38
3.3a	Rigidity neck	1.39	–0.55	0.79	2.22	3.72	1.55	1.84
3.3b	Rigidity RUE	1.28	–1.23	0.47	2.40	4.42	1.52	2.44
3.3c	Rigidity LUE	1.40	–1.05	0.47	2.20	4.19	1.45	2.26
3.3d	Rigidity RLE	1.51	–0.61	0.68	2.19	3.80	1.52	1.91
3.3e	Rigidity LLE	1.58	–0.61	0.61	2.06	3.65	1.43	1.84
3.4a	Finger tapping—right hand	1.92	–1.30	0.13	1.37	2.89	0.77	1.78
3.4b	Finger tapping—left hand	2.16	–1.23	–0.02	1.13	2.63	0.63	1.65
3.5a	Hand movements—right hand	2.09	–1.01	0.35	1.59	3.13	1.02	1.77
3.5b	Hand movements—left hand	2.27	–1.01	0.22	1.41	2.77	0.85	1.62
3.6a	Pronation supination—right hand	1.98	–1.01	0.33	1.56	2.94	0.96	1.69
3.6b	Pronation supination—left hand	2.03	–1.08	0.16	1.31	2.67	0.77	1.60
3.7a	Toe tapping—right foot	2.14	–1.02	0.28	1.38	2.61	0.81	1.55
3.7b	Toe tapping—left foot	2.17	–1.09	0.09	1.14	2.40	0.64	1.49
3.8a	Leg agility—right leg	2.40	–0.73	0.50	1.55	2.67	1.00	1.45
3.8b	Leg agility—left leg	2.49	–0.80	0.33	1.33	2.48	0.84	1.40
3.9	Arising from chair	1.89	0.05	1.14	1.80	2.55	1.39	1.06
3.1	Gait	1.85	–1.27	0.41	1.58	2.70	0.86	1.70
3.11	Freezing of gait	1.52	0.64	1.52	2.38	3.04	1.90	1.04
3.12	Postural stability	1.52	–0.24	0.73	1.34	2.79	1.16	1.27
3.13	Posture	1.58	–1.25	0.42	1.77	3.16	1.03	1.88
3.14	Global spontaneity of movement	2.21	–1.45	0.04	1.23	2.63	0.61	1.74
3.15a	Postural tremor—right hand	0.52	0.60	4.46	7.46	12.09	6.15	4.85
3.15b	Postural tremor—left hand	0.58	0.52	3.96	7.18	10.86	5.63	4.42
3.16a	Kinetic tremor—right hand	0.50	1.23	5.22	9.47	13.19	7.28	5.18
3.16b	Kinetic tremor—left hand	0.52	0.97	4.57	9.04	13.07	6.91	5.27
3.17a	Rest tremor—RUE	0.44	1.00	3.48	6.58	11.55	5.65	4.55
3.17b	Rest tremor—LUE	0.54	0.91	3.04	5.88	9.75	4.90	3.82
3.17c	Rest tremor—RLE	0.47	3.06	5.53	9.02	14.29	7.98	4.87
3.17d	Rest tremor—LLE	0.59	2.75	4.81	7.70	12.19	6.86	4.09
3.17e	Rest tremor—Lip	0.65	3.36	5.49	8.11	9.79	6.69	2.84
3.18	Constancy of rest tremor	0.39	–0.69	1.88	4.10	6.61	2.98	3.12

RUE, right upper extremity; LUE, left upper extremity; RLE, right lower extremity; LLE, left lower extremity; SD, standard deviation.

TABLE 3. Item factor loading for all MDS-UPDRS Part 3 Items in the multidimensional model

Items		Factor 1	Factor 2
3.1	Speech	0.653	
3.2	Facial expression	0.644	
3.3a	Rigidity neck	0.616	
3.3b	Rigidity RUE	0.556	
3.3c	Rigidity LUE	0.606	
3.3d	Rigidity RLE	0.644	
3.3e	Rigidity LLE	0.671	
3.4a	Finger tapping—right hand	0.718	
3.4b	Finger tapping—left hand	0.774	
3.5a	Hand movements—right hand	0.748	
3.5b	Hand movements—left hand	0.789	
3.6a	Pronation supination—right hand	0.732	
3.6b	Pronation supination—left hand	0.757	
3.7a	Toe tapping—right foot	0.775	
3.7b	Toe tapping—left foot	0.796	
3.8a	Leg agility—right leg	0.807	
3.8b	Leg agility—left leg	0.835	
3.9	Arising from chair	0.754	
3.10	Gait	0.751	
3.11	Freezing of gait	0.692	
3.12	Postural stability	0.690	
3.13	Posture	0.688	
3.14	Global spontaneity of movement	0.788	
3.15a	Postural tremor—right hand		–0.739
3.15b	Postural tremor—left hand		–0.682
3.16a	Kinetic tremor—right hand		–0.670
3.16b	Kinetic tremor—left hand		–0.593
3.17a	Rest tremor—RUE		–0.863
3.17b	Rest tremor—LUE		–0.812
3.17c	Rest tremor—RLE		–0.791
3.17d	Rest tremor—LLE		–0.771
3.17e	Rest tremor—Lip		–0.711
3.18	Constancy of rest tremor		–0.897

Factor loading ≤0.30 are not displayed.
RUE, right upper extremity; LUE, left upper extremity; RLE, right lower extremity; LLE, left lower extremity.

with the 23 nontremor items loading on one factor (minimal factor loading: 0.541) and the 10 tremor items loading on the second factor (minimal factor loading = 0.441). No items demonstrated dual-factor loadings ≥0.30 (Table 3).

When we investigated the IRT discrimination and location parameters for the 10 tremor items from Part 3 of the MDS-UPDRS considered separately from the 23 nontremor items, the discrimination parameters for the tremor items increased from low to high or very high (mean = 2.13 ± 0.49). The location parameters revealed an improvement in range, representing a reasonable relationship between item score and the latent trait of overall parkinsonian severity (location parameter means ranging from –0.12 to 5.29; Table 4). The ICCs for this analysis can be found in the Supporting Information (see Supporting Information Fig. S1B).

Estimation of the scores of the 10 tremor items from both the unidimensional model and the multidimensional model is presented in mirror plots and scatter plots, and suggest that the observed mean proportion of response in each category of each tremor item was sufficiently estimated by both the uni- and multidimensional models and the standard errors were very small because of the large sample size (see Supporting Information Figs. S2 and S3). Figure 1 compares the goodness of fit of the item 3.17a (Rest tremor amplitude-RUE) from the unidimensional model (upper panels) and the multidimensional model (lower panels). It suggests that the multidimensional model better fits the observed data of item 3.17a than the unidimensional model. The same conclusion holds for other tremor items (see Supporting Information Figs. S4 and S5).

Discussion

Since its introduction in 2007, the MDS-UPDRS has undergone extensive clinimetric examination with very satisfactory results analyzed under the CTT principles.[4] Given the wide and complex scope of psychometric

TABLE 4. Discrimination (Discrim) and item location parameters for the 10 tremor items of MDS-UPDRS Part 3 considered separately

			Item Location					
Items		Discrim	0 to 1	1 to 2	2 to 3	3 to 4	Mean	SD
3.15a	Postural tremor—right hand	1.95	0.28	1.68	2.65	4.03	2.16	1.58
3.15b	Postural tremor—left hand	1.86	0.27	1.70	2.89	4.17	2.26	1.67
3.16a	Kinetic tremor—right hand	1.66	0.53	2.04	3.48	4.68	2.68	1.80
3.16b	Kinetic tremor—left hand	1.45	0.47	2.03	3.78	5.29	2.89	2.09
3.17a	Rest tremor—RUE	2.53	0.37	1.08	1.85	2.95	1.56	1.10
3.17b	Rest tremor—LUE	2.57	0.40	1.14	2.00	3.02	1.64	1.13
3.17c	Rest tremor—RLE	2.07	1.10	1.86	2.81	4.17	2.49	1.32
3.17d	Rest tremor—LLE	2.20	1.16	1.89	2.82	4.19	2.52	1.31
3.17e	Rest tremor—lip	1.86	1.61	2.48	3.48	4.11	2.92	1.10
3.18	Constancy of rest tremor	3.12	–0.12	0.58	1.10	1.59	0.79	0.73

RUE, right upper extremity; LUE, left upper extremity; RLE, right lower extremity; LLE, left lower extremity; SD, standard deviation.

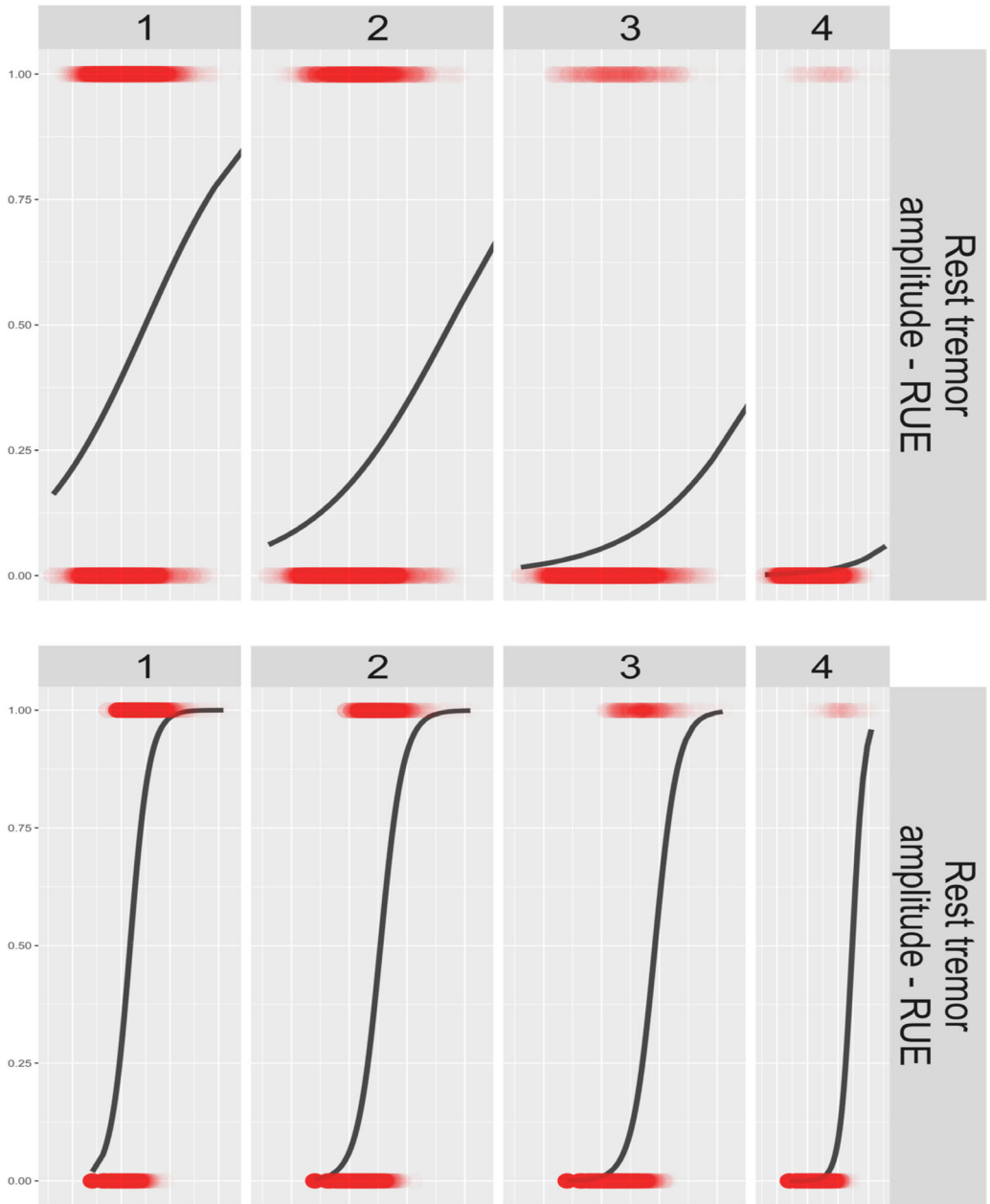

FIG. 1. Plots display the ICC fits of the cumulative probabilities of the item 3.17a (Rest tremor amplitude-RUE) from the unidimensional model (upper panels) and the multidimensional model (lower panels). Red dots indicate the observed mean response proportions. The panels with "1" (left most column) show the values with scores 0 and ≥ 1; the panels with "2" (second column) show values with scores ≤1 and ≥2; the panels with "3" (third column) show values with scores ≤2 and ≥3; and the panels with "4" (fourth column) show values with scores ≤3 and 4. RUE, right upper extremity. [Color figure can be viewed at wileyonlinelibrary.com]

approaches outside of CTT, each allowing new analyses and the application of statistical tests anchored in different models, we have utilized the polytomous IRT technique, the graded response model, to focus specifically on the most widely reference portion of the scale, the Motor Examination.[1] We used this analytical model to assess the contribution of items that, in their multiple responses, assess motor severity through accurate measurements grounded on the same overall metric, the latent-trait model.[13] If there is a single and unifying disease severity that is captured by all the items, a high metric severity value can be tallied. If, on the other hand, individual or clusters of items separate from the rest of the scale, two different measurements of disease severity are likely embedded in the overall measurement, and separating them for independent consideration may provide more accurate information on motor severity.[23,24]

The specific issue of interest in our study was dissecting the Motor Examination in an IRT framework in

order to define, in a cross-section cohort of patients expressing the full gamut of disease severity, the items and responses calibrated to a sample-independent latent model of PD. Our finding, that tremor and nontremor items distinctly separate across the gamut of the disease, provides a strong rationale for viewing disease progression and therapeutic treatments in the light of separate effects on these two indices or groupings. Disease severity, defined, by a summary score that coalesces all Part 3 items is, in fact, balancing deficits in two separate domains so that 2 patients with the same total score may reflect equal deficits across all items or a very different phenotype with disproportionate severities of tremor versus nontremor. Likewise, in a longitudinal assessment, a stable total severity score may accurately indicate clinical stability, but may also hide salient information on differential progression or treatment response for tremor or nontremor impairments. These findings indicate the need to consider Part 3 as a total summary inventory, but the analytical approaches in prospective studies may need to consider tremor items and nontremor items separately.

As background, previous CTT examinations of the Motor Examination (Part 3) demonstrated an invariant single-factor construct,[5,7] independent of the other Parts, and within this single factor, seven subgroupings covering axial and appendicular bradykinesia, rigidity, gait, posture and action, and postural and resting tremor. Comparatively, previous IRT examinations of the MDS-UPDRS have focused on longitudinal analyses in a cohort of patients followed from <3 years of onset for up to 4 years of follow-up. These studies have combined items from the separate Parts of the MDS-UPDRS in nontraditional scoring methods, often combining examiner administered items of nonmotor functional impact and motor examination with patient reported measures of nonmotor and motor functional impact on experiences of daily living. For this study, we analyzed the MDS-UPDRS Part 3 Motor Examination items from a large worldwide representative sample of PD patients expressing the full gamut of disease severity, using the traditional scoring of the examiner administered motor examination (Part 3). We used an IRT model because of its ability to assess the relationships between item scores and a sample-invariant estimate of overall parkinsonian severity using latent-trait modeling. We applied the graded response model to estimate the discrimination and location parameters for all polytomous items of Part 3. Items measuring rigidity, bradykinesia, gait, and posture presented satisfactory parameters of discrimination and location, documenting a high or very high relationship to motor severity with appropriate scaling to the latent measure. However, the 10 items measuring tremor presented unsatisfactory results when included with those items assessing bradykinesia, rigidity, gait, and posture. The low to very low discrimination, along with the poor location parameters of the tremor items, demonstrates an inadequate mapping of these items to the overall latent model of parkinsonian motor severity when combined with the other 23 items. This finding suggests that tremor items add "noise" in the total score of Part 3 for modeling overall parkinsonian severity when combined with the items assessing bradykinesia, rigidity, gait, and posture. Likewise, these latter items add noise to the overall severity of parkinsonism determined by tremor items considered by themselves.

When we tested this theory by comparing the unidimensional model to a multidimensional model, we found clear superiority for the two-dimensional model. The multidimensional model clearly separated the 23 items measuring rigidity, bradykinesia, gait, and posture from the 10 items measuring tremor with no cross-loadings across factors. When the 10 tremor items were analyzed separately, there was a considerable improvement in discrimination, from Fabrozzi-based rankings of very low or low to high or very high, and a normalization of the location parameters to a reasonable range.[21] These results are consistent with the improved accuracy of responses on the latent characteristic continuum confirmed by the multidimensional model. Taken together, these results suggest that the sensitivity of Part 3 is categorically improved by considering the tremor items separately from those assessing bradykinesia, rigidity, and posture.

As mentioned before, a study that used the IRT's Rasch model to assess psychometric aspects of MDS-UPDRS Parts 2 and 3 in early-stage PD patients showed psychometric limitations that compromise the accuracy of motor symptom measurements and their impact in early PD. The results for Part 3 identified measurement weaknesses for various items, including facial expression, speech, posture, gait, and tremor, for PD patients in the early disease course.[18] Our results differ because, in addition to using the IRT's graded-response model, which is better suited for the MDS-UPDRS analysis, we extend the findings across the full gamut of motor severity of PD patients and identify items that evaluate tremor as specifically more sensitive when analyzed separately from nontremor items.

Recognition of this division is already established in empirical definitions of the PD subtypes of tremor-dominant versus posture instability and gait disorder.[2,25] From a clinical point of view, clinicians may agree that tremor behaves differently from bradykinesia and rigidity. In the clinic, we send subjects with severe tremor, often unresponsive to medication, for DBS even though for most patients, we expect DBS to do no more than levodopa at its best efficacy in terms of item-by-item improvement in MDS-UPDRS Part 3 scores. Treatment and pathophysiological differences between tremor and nontremor symptoms are also observed.

Marked tremor is often not responsive to dopaminergic therapy and does not appear to be related to dopaminergic deficiency in early PD.[26] Cellular recordings demonstrate tremor-related extrabasal ganglia activity in the ventralis intermedius nucleus of the thalamus.[27]

The definition of the domain of a scale is one of the first phases of its development and is based on the most current knowledge about the construct that is being studied.[23,24] Our results demonstrate that within the domain of PD motor severity, assessments of tremor may be best considered separately from assessments of bradykinesia, rigidity, posture, and gait. Practically speaking, if two distinct behaviors underpin one final disease severity, what is the clearest and most interpretable measure of this dichotomous reality? We suggest that if the MDS-UPDRS were to be revised again, the single Motor Examination score may be revised to tremor and nontremor divisions.

For current applications, this cross-sectional finding should be replicated in other dataset by separating the tremor items from the nontremor items to provide more sensitive measures of the underlying trait of PD motor severity related to each set of items. The increased sensitivity of separating tremor and nontremor measures may improve the outcome of clinical trials where the total Part 3 Motor Examination score is used as the primary outcome measure of efficacy. As such, a treatment with little effect on nontremor aspects of PD, but with substantive antitremor efficacy, could be identified easily and with a reduced sample size. Agents that positively impact the tremor domain, but actually aggravate the nontremor domain, could be potentially modified biochemically to eliminate the problematic moiety impacting nontremor function, but retain the antitremor impact. Differential impact of placebo treatment on these two aspects of PD severity has never been studied, but is a subject of pertinent interest for study design and power calculation projections. To assess this potential increased sensitivity, application of this multimodal conceptualization of MDS-UPDRS Part 3 scores in longitudinal studies will need to be examined.

Despite the limitations inherent in the cross-sectional design of our study, and the limited enrollment of severe cases (with only 2% of the patients on H & Y stage 5), our results demonstrate that the focused use of appropriate statistical tools can assist in uncovering the item-level associations to a latent measure of PD severity. The lack of extensive disease history information, such as current medications and dosages, limit our analyses to the manifestation of the motor impairment regardless of treatment regimens. The results do not refute the findings of the CTT analyses of Part 3 demonstrating a single factor structure. Rather, our findings increase our understanding of the underlying relationship between scale items and the sample-independent measurement of PD motor severity, whereby tremor items appear to have a separate relationship to underlying PD severity from items assessing bradykinesia, rigidity, posture, and gait. ■

Acknowledgments: The Rush Parkinson's Disease and Movement Disorders Program is a designated Clinical Center of Excellence supported by the Parkinson Foundation.

References

1. Furr RM. Psychometrics: An Introduction [Internet], 3rd ed. Thousand Oaks, CA: Sage Pubications, Inc.; 2018.
2. Jankovic J, McDermott M, Carter J, et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. Neurology 1990;40:1529–1529.
3. Ramaker C, Marinus J, Stiggelbout AM, van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord 2002;17:867–876.
4. Goetz CG. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. Mov Disord 2003;18:738–750.
5. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008;23:2129–2170.
6. Goetz CG, Stebbins GT, Tilley BC. Calibration of Unified Parkinson's Disease Rating Scale scores to Movement Disorder Society-Unified Parkinson's Disease Rating Scale scores. Mov Disord 2012;27:1239–1242.
7. Goetz CG, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): process, format, and clinimetric testing plan. Mov Disord 2007;22:41–47.
8. Gallagher DA, Goetz CG, Stebbins G, Lees AJ, Schrag A. Validation of the MDS-UPDRS Part I for nonmotor symptoms in Parkinson's disease. Mov Disord 2012;27:79–83.
9. DeVellis RF. Classical test theory. Med Care 2006;44(Suppl. 3): S50–S59.
10. Nunnally JC, Bernstein IH. Psychometric Theory, 3rd ed. New York, NY: McGraw-Hill; 1994.
11. Kean J, Reilly J. Item response theory. Lang Test 1985;2:197.
12. Forero CG, Maydeu-Olivares A. Estimation of IRT graded response models: limited versus full information methods. Psychol Methods 2009;14:275–299.
13. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res 2007;16(Suppl. 1):5–18.
14. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care 2000;38(9 Suppl.):II28–II42.
15. Buatois S, Retout S, Frey N, Ueckert S. Item response theory as an efficient tool to describe a heterogeneous clinical rating scale in de novo idiopathic Parkinson's disease patients. Pharm Res 2017;34: 2109–2118.
16. Gottipati G, Karlsson MO, Plan EL. Modeling a composite score in Parkinson's disease using item response theory. AAPS J 2017;19: 837–845.
17. Gottipati G, Berges AC, Yang S, Chen C, Karlsson MO, Plan EL. Item response model adaptation for analyzing data from different versions of Parkinson's disease rating scales. Pharm Res 2019; 36:135.
18. Regnault A, Boroojerdi B, Meunier J, Bani M, Morel T, Cano S. 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: 1927–1936.
19. Goetz CG, Stebbins GT, Wang L, LaPelle NR, Luo S, Tilley BC. IPMDS-sponsored scale translation program: process, format, and clinimetric testing plan for the MDS-UPDRS and UDysRS. Mov Disord Clin Pract 2014;1:97–101.

20. R Core Team. R: a language and environmment for statiscal computing. Vienna: R Foundation for Statistical Computing; 2019.

21. Fabozzi FJ, Focardi SM, Rachev ST, Arshanapalli BG, Hoechstoetter M. The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications. Hoboken, NJ: John Wiley & Sons; 2014.

22. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. J Stat Softw 2012;48:1–29.

23. Benson J, Clark F. A guide for instrument development and validation. Am J Occup Ther 1982;36:789–800.

24. Stebbins GT. Clinical rating scale development. In: Sampaio C, Goetz CG, Schrag A, eds. Rating Scales in Parkinson's Disease: Clinical Practice and Research. New York, NY: Oxford University Press; 2012:5.

25. Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify tremor dominant and postural instability/gait difficulty groups with the Movement Disorder Society Unified Parkinson's Disease Rating Scale: comparison with the Unified Parkinson's Disease Rating Scale. Mov Disord 2013;28:668–670.

26. Lee JY, Lao-Kaim NP, Pasquini J, Deuschl G, Pavese N, Piccini P. Pallidal dopaminergic denervation and rest tremor in early Parkinson's disease: PPMI cohort analysis. Parkinsonism Relat Disord 2018;51:101–104.

27. Hallett M. Parkinson's disease tremor: pathophysiology. Parkinsonism Relat Disord 2012;18(Suppl. 1):S85–S86.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

SGML and CITI Use Only
DO NOT PRINT

Author Roles

(1) Research Project: A. Conception, B. Organization, C. Execution; (2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique; (3) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique.

M.H.S.T.: 1A, 1B, 1C, 2B, 3A
C.G.G.: 1B, 2C, 3C
S.L.: 1B, 2C, 3B
D.C.: 2B, 3C
G.T.S.: 1A, 1B, 1C, 2A, 2B, 3A

Financial Disclosures

C.G.G. has received funding to Rush University Medical Center from the NIH, Department of Defense, and Michael J. Fox Foundation for research conducted by Dr. Goetz. During the reporting time, Dr. Goetz directed the Rush Parkinson's Disease Research Center supported by the Parkinson's Foundation, and some of these funds supported Dr. Goetz's salary as well as his research efforts. Dr. Goetz has received a presidential stipend from the International Parkinson and Movement Disorder Society paid to Rush University Medical Center as part of Dr. Goetz's salary, faculty stipends from the International Parkinson and Movement Disorder Society and the American Academy of Neurology, and a guest professorship honorarium provided by the NorthShore University Health System. Dr. Goetz has received royalties from Oxford University Press and Wolters Kluwer Publishers and has received a salary from Rush University Medical Center. S.L. has consulted for and held advisory board membership with honoraria with the National Institutes of Health, CHDI Management, Inc., T3D therapeutics, Kashiv BioSciences, and MyMee Inc, GuidePoint; grant and research support from the National Institutes of Health, CHDI, and International Parkinson and Movement Disorder Society; has received honoraria from the Academy of Mathematics and Systems Science of China, East China Normal University, and Central South University of China and has received a salary from Duke University. D.C. has received a salary from Duke University. G.T.S. has consulted for and held advisory board membership with honoraria from Acadia, Pharmaceuticals, Adamas Pharmaceuticals, Inc., Biogen, Inc., Ceregene, Inc., CHDI Management, Inc., Cleveland Clinic Foundation, Ingenix Pharmaceutical Services (i3 Research), MedGenesis Therapeutix, Inc., Neurocrine Biosciences, Inc., Pfizer, Inc., Tools-4-Patients, Ultragenyx, Inc., and Sunshine Care Foundation; has received grant and research support from the National Institutes of Health, Department of Defense, Michael J. Fox Foundation for Parkinson's Research, Dystonia Coalition, CHDI, Cleveland Clinic Foundation, International Parkinson and Movement Disorder Society, and CBD Solutions; has received honoraria from the International Parkinson and Movement Disorder Society, American Academy of Neurology, Michael J. Fox Foundation for Parkinson's Research, U.S. Food and Drug Administration, National Institutes of Health, and Alzheimer's Association and has received a salary from Rush University Medical Center.