Published in final edited form as:

Mov Disord. 2015 October ; 30(12): 1632–1638. doi:10.1002/mds.26153.

Handling Missing Values in the MDS-UPDRS

Christopher G. Goetz, MD^{1,*}, Sheng Luo, PhD², Lu Wang, MS², Barbara C. Tilley, PhD², Nancy R. LaPelle, PhD³, and Glenn T. Stebbins, PhD¹

¹Department of Neurological Sciences, Rush University Medical Center, Chicago IL, USA

²Division of Biostatistics, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA

³Division of Preventive and Behavioral Medicine, University of Massachusetts, Worcester, MA, USA

Abstract

This study was undertaken to define the number of missing values permissible to render valid total scores for each Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) part. To handle missing values, imputation strategies serve as guidelines to reject an incomplete rating or create a surrogate score. We tested a rigorous, scale-specific, data-based approach to handling missing values for the MDS-UPDRS. From two large MDS-UPDRS datasets, we sequentially deleted item scores, either consistently (same items) or randomly (different items) across all subjects. Lin's Concordance Correlation Coefficient (CCC) compared scores calculated without missing values with prorated scores based on sequentially increasing missing values. The maximal number of missing values retaining a CCC greater than 0.95 determined the threshold for rendering a valid prorated score. A second confirmatory sample was selected from the MDS-UPDRS international translation program. To provide valid part scores applicable across all Hoehn and Yahr (H&Y) stages when the same items are consistently missing, one missing item from Part I, one from Part II, three from Part III, but none from Part IV can be allowed. To provide valid part scores applicable across all H&Y stages when random item entries are missing, one missing item from Part I, two from Part II, seven from Part III, but none from Part IV can be allowed. All cutoff values were confirmed in the validation sample. These analyses are useful for constructing valid surrogate part scores for MDS-UPDRS when missing items fall within the identified threshold and give scientific justification for rejecting partially completed ratings that fall below the threshold.

Keywords

Parkinson's disease; rating scales; MDS-UPDRS; missing values; Lin's correlation coefficient

*Correspondence to: Dr. Christopher Goetz, 1725 West Harrison Street, Suite 755, Chicago, IL 60612-3864, cgoetz@rush.edu.

Relevant conflicts of interest/financial disclosures: The authors received compensation from the Movement Disorder Society for the management of this program.

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

Supporting Data

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

For the application of any rating scale in both clinical practice and research programs, when a value for an item in the scale is missing, the handling of that problem, either by rejecting the rating or creating a surrogate score, is challenging. Different methods have been used to salvage a score when item entries are missing. These include “last value carried forward,” a problematic technique because of bias in a disease that is chronically progressing, and simple or multiple imputation strategies, including the prorated score, a technique that can be applied when at least half of the items have complete data and when the items are scaled using only one metric.^{1–5} The prorated score is derived by taking the mean of the observed scores and substituting that value for all of the missing values in the scale. Because the MDS-UPDRS has a consistent metric of 0 to 4 ratings across all items, we sought to test the prorated method using a data-based approach and rigorous threshold for handling missing values in this scale. With access to two large datasets, we investigated the maximum number of missing values allowed while maintaining a near perfect matching with the complete-value score. We used the original validation sample from the English MDS-UPDRS^{6,7} to model the maximal number of missing values for each part that would still allow a valid total score to be calculated. We then confirmed our findings using the independently derived validation data from the MDS-UPDRS international translation program.⁸

Methods

In the original clinimetric validation analysis of the MDS-UPDRS,^{6,7} we obtained complete scores for 842 English-speaking PD patients. All items that make up the total score were scored on a scale of 0 to 4. The men and women, aged 31 to 98, represented a wide variety of races and ethnicities.⁶ Whereas the sample included all Hoehn and Yahr stages (H&Y), the distribution was not equal: stage 1–2, N = 516; stage 3, N = 166; and Stage 4–5, N = 160. To test whether our estimates differed specifically according to disease severity and to control for possible confounds because of sample size differences, we used all scores from stages 4–5 and randomly selected MDS-UPDRS scores from the larger groups (H&Y 1–2 and H&Y 3) so that each of the three samples had 160 individuals (final N = 480) for this analysis.

The MDS-UPDRS is designed as four separate parts, with summary scores for each to provide an overall severity measure of a given aspect of Parkinson’s disease (part I, non-motor experiences of daily living; part II, motor experiences of daily living; part III, motor examination; part IV, motor complications).^{6,7} Our previous work had shown that the parts had to be analyzed separately.^{6,7} To estimate the number of permissible missing values for each part, we first computed the score as the sum of each patient’s individual item scores for each part. This score for each part with complete data was used as the gold standard with which the prorated scores based on purposeful deletion of item scores (missing values) could be compared. In this process, the prorated score was calculated as the sum of the available scores multiplied by the number of total items in the complete part of the MDS-UPDRS, and this result was divided by the number of items with actual scores. Comparison between the missing value-based prorated score and the complete score was evaluated using Lin’s concordance correlation coefficient (CCC).⁹ The CCC measures the exact matching between values (scores generated with all data vs. prorated score with missing values). We set our critical CCC level at 0.95 or greater, interpreted as near-perfect agreement between the

missing value–based and the full data–based scores.⁹ To accommodate different clinical situations, we applied two approaches to calculate the missing values' prorated scores. First, we systematically omitted a consistent item score (eg, cognitive impairment) for all cases from a given part of the MDS-UPDRS, starting with one missing item and extending up to the maximal number of missing values that still maintained a CCC of 0.95 or greater for all combinations of consistently missing items. This approach modeled the clinical situation of a consistent elimination of items in each part across all patients, as might occur with the practical constraints of field work or telemedicine, in which certain questions could be considered too sensitive to apply outside the office setting, where safety concerns (postural stability) preclude inclusion, or where an item or items are inadvertently excluded in the questionnaire design. We refer to this approach of omitting items as consistently missing.

In the second approach, we omitted a given number of item scores randomly selected across cases (eg, cognition from one patient, hallucinations and psychosis from another), starting with one selected omission and extending up to the maximal number of missing values that retained a CCC of 0.95 or higher. This analysis mimicked a clinical trial or day-to-day clinic encounter in which raters mistakenly overlook items, so that the final sample has missing values that vary across raters and sites. We refer to this approach of omitting items as randomly missing. For the randomly missing approach, the CCC was calculated with 1,000 randomly selected replications of missing items in all combinations for each MDS-UPDRS part.

Because the number of part III items (33) is much higher than the number constituting the other parts (13 in part I, 13 in part II, and 6 in part IV), and because part III has specific right side, left side, and midline assessments, we also examined the part III items with the same techniques, deleting items consistently and randomly, but selecting the deletions from within the item sub-categories:

Midline: speech, facial expression, neck rigidity, arising from chair, gait, freezing, postural stability, posture, global spontaneity, rest tremor of lip/jaw

Right Side: upper extremity rigidity, lower extremity rigidity, finger taps, hand movements, pronation–supination, toe taps, leg agility, postural tremor, kinetic tremor, upper extremity rest tremor, lower extremity rest tremor

Left Side: upper extremity rigidity, lower extremity rigidity, finger taps, hand movements, pronation–supination, toe taps, leg agility, postural tremor, kinetic tremor, upper extremity rest tremor, lower extremity rest tremor

For constancy of rest tremor (item 3.18), we examined all five rest tremor scores and wherever the highest rating resided, we assigned this item to that location (right, left, or midline). For example, if the highest rest tremor score was right upper extremity (RUE), we assigned Constancy of Rest Tremor for that patient to the Right Side domain. If RUE and left upper extremity (LUE) both had the same rating and were higher than all other scores for rest tremor, we assigned constancy of rest tremor to both right and left. If the highest value was seen on right, left, and jaw/lip, we assigned constancy of rest tremor to all three categories.

Our prespecified analytic plan also included a confirmatory step using the samples from all official non-English versions of the MDS-UPDRS.^{8,10,11} Following the same methodology described earlier, we assembled all cases that were part of the completed official translations (Chinese, Japanese, Estonian, French, Hungarian, Russian, Spanish, Italian, German, and Slovakian; total sample size of 3,784 PD cases). We divided these entries by H&Y stages (stages 1–2, 3, and 4–5). The smallest sample (stage 4–5) was 380 cases and, to equalize the sample sizes of all groups, we randomly selected MDS-UPDRS scores from stage 1–2 and 3 so that each of the three samples had 380 individuals (final N = 1,140) for the confirmatory analysis. We used the number of acceptable missing values from the original analysis and tested whether the CCC remained at 0.95 or greater for each MDS-UPDRS part in the validation group. As before, we tested the number of allowable missing values for both consistently and randomly missing items.

Results

Consistent Deletion of the Same Item Across All Patients

When given items were consistently dropped from part I of the scale (13 items total; Table 1), the minimum CCC remained 0.95 or greater across all H&Y stages for one missing item. With two consistently deleted items, the minimum CCC fell below the 0.95 or greater threshold and did not allow an adequate prorated score to match the actual part I score generated by complete data. Similarly, for part II (13 items total), when any single item was consistently deleted from the data set, the minimum CCC remained 0.95 or greater for all H&Y groups, but an accurate prorated score could not be obtained if two items were consistently absent. For part III (33 items total), across all H&Y stages, three items could be missing, with the minimum CCC remaining 0.95 or greater. In the case of H&Y stages 1–2, for part III, four consistently missing values could be dropped with a maintained accurate score calculation. In contrast, for part IV, all data were essential to include, and even one missed value led to CCC of 0.95 or greater.

Random Deletion of Items

When item scores were deleted at random within a given part, for part I, a total part I score could be validly calculated across all H&Y stages if only one item per patient was missing. In the case of H&Y stages 1–2 and 3, two items could be missing, but in stages 4–5, only one was acceptable with a CCC of 0.95 or greater. For part II, across all HY stages, a total part II score could still be validly calculated if there were two randomly deleted items (Table 2). In the case of H&Y stages 1–2, three items could be missing and the CCC still remained above threshold, but in stages 3 and 4–5, only two missing values could be tolerated to provide a prorated score that accurately compared with the part II score generated with a full data set. For part III, across all H&Y stages, up to and including seven items could be missing at random with an accurate total part III score still validly calculable. In the case of H&Y stages 1–2 and 3, nine items could be missing, but in stages 4–5, only seven missing values were tolerated and still provided sufficient information, so that the prorated score calculation closely matched the actual part III score based on full data. For part IV, across all H&Y stages, a total part IV score could not be accurately rendered when any randomly missing item was deleted. In the case of H&Y stages 1–2, one item could be missing, but in

all other stages, any missing values precluded the calculation of a prorated score that fit with the score generated from the full data set.

Summary of Permissible Missing Values Consistently or Randomly Deleted (Table 3)

In summary, to provide a valid part score that applies across all H&Y stages, the MDS-UPDRS can afford one consistently missing item's value from part I, one from part II, three from part III, but none from part IV (Table 3). In contrast, to provide a valid part score that applies across all H&Y stages, the MDS-UPDRS can afford, in a given patient, one randomly missing item's value from part I, two from part II, seven from part III, but none from part IV.

Further Analysis of Part III

When the 33 Part III items were divided into right side items, left side items and midline items (see Methods), these analyses provided fully compatible results to the prior estimates for the number of allowable missing values to those calculated for all H&Y stages considered together (Supplemental Data Table 1). If two items assessing right side, two items assessing left side function, or two items assessing midline function were consistently deleted from part III of the MDS-UPDRS, the prorated score calculated by the remaining data accurately generated scores reflective of the same subsection standard score based on the full data set. Further deletions to three missing values in any subsection, however, dropped the minimum CCC below 0.95 for that subsection, indicating that the remaining data did not provide a valid prorated score.

When the same analysis was performed with randomly deleted items from within the three part III subsection categories, for all three subsections, if three items from a given section were randomly deleted, the CCC for that subsection remains 0.95 or greater, indicating that a valid score for that section could still be calculated. Further deletions, however, dropped the CCC below threshold, indicating that the MDS-UPDRS loses validity for these sections when more than three missing values occur.

Confirmatory Analysis Using Data Sets From Official MDS-UPDRS Non-English Editions

The confirmatory analysis using the larger sample size from the combined non-English official versions of the MDS-UPDRS (see Methods) confirmed our original findings (Table 4). The number of allowable missing items identified as applicable across all H&Y stages in the original sample was validated for both consistently and randomly deleted items when we tested these values in the combined non-English sample.

To guide clinicians and researchers on how to apply these calculations in a practical setting, several examples are provided (Fig. 1).

Discussion

Missing data in clinical rating scales are problematic, because assessments are usually time-locked to an office visit and cannot be retrospectively inserted with reliability. Several methods have been developed to handle the problem of missing data, ranging from setting an

arbitrary cutoff for the maximum number of missing values to complex imputation methods.¹⁻⁵ However, such strategies serve as guidelines, and validity of any method for a given scale can only be established by testing the techniques against full data sets. Based on a long history of using rating scales to assess Parkinson's disease and with experience in dealing with data sets with missing values, we approached this analysis from two perspectives: first, we recognize that the MDS-UPDRS can be difficult to perform in certain circumstances, such as in field work where untrained professionals may be gathering data, and such items as postural stability (3.11) are not considered safe to perform. Likewise, protocols with intravenous infusions in one arm may compromise the ability to rate parkinsonian features on the encumbered side of the body. In these instances, items may be consistently lost from the data set, creating a dilemma for computing the part III score; second, and a more common problem for practice and clinical trials, items may be inadvertently left unrated, with the specific missing items varying from patient to patient. This problem can affect part Ia, III, and IV, where the investigator may forget to ask a question or evaluate a question, but it also can occur in parts Ib and II if the patient overlooks an item on the questionnaire. In this study, we have identified a clear threshold for how many missing items can exist in each part of the MDS-UPDRS and still allow a solid calculation of a total part score for both of these clinical situations. Equally important, we identify the threshold for needing to reject MDS-UPDRS data when too many items are missing.

We selected the Lin's CCC, because this test measures the exact matching between two values, rather than the relative order of values, as measured by Pearson's Product-Moment Correlation Coefficient (r) or the ranking of values, as measured by Spearman's Rank Order Correlation Coefficient (ρ).^{2,9} We purposely set our critical CCC at a very high level of agreement between the missing value-based and the gold-standard scores calculated with full data, allowing a maximal mean difference of less than 10%. The method allows for the accurate calculation of total part scores, but it does not allow individual item scores to be generated, except in the case of a single missed value. This focus on part total scores is clinically justified, because patient monitoring is based on total part scores, not individual items. The confirmatory analysis based on the large data set derived from the non-English MDS-UPDRS validation program was uniformly successful in demonstrating that our original calculations were applicable to a fully independent data set.

With the exception of part III, the number of missing items permissible for generating a valid total score was small, suggesting that for parts I, II, and IV, each item is assessing an important feature of the disease without duplicative data. In contrast, for part III, the number of allowable missing items was surprisingly high and, even with our sub-analysis of right side, left side, and midline missing values, allowable missing items exceeded those seen with the other parts. The original team constructing the MDS-UPDRS focused its effort on providing comprehensive, but maximally efficient assessments, but these analyses suggest that future work could focus on the feasibility of a shorter part III with item reductions.

As a strength of our study, we have based our analyses on large and uniformly complete data sets of MDS-UPDRS scores, and we then purposefully dropped item values in different patterns to assess their impact. Our recommendations therefore are based on actual data and

not estimates or empiric guidelines. We have constructed our analyses to mimic the two clinically pertinent situations: first, where given items are consistently missed, and second, where items are randomly lost because of error. As limitations, we acknowledge that our method allows a surrogate total part score, but individual items cannot be generated unless only one value is missing. Conversely, the MDS-UPDRS was not constructed for individual item monitoring, and clinical trials and care focus on total part scores.^{6,7} Furthermore, although we selected large datasets that purposefully enrolled patients across severity stages, we cannot ensure that they are representative of all future datasets. Finally, whereas we tested our findings in two independent datasets with full confirmation across all MDS-UPDRS parts, there may be individuals with unusual item scores that do not follow these population patterns for unknown reasons, and therefore the prorated score will not be accurately reflective of the missing value in these individuals. Because these issues are implicit to any imputation technique, future emphasis needs to be placed on techniques to eliminate missing values on the part of raters and patients. Although the methods we used potentially could be applicable to determining the number of permissible missing items for other instruments constructed similarly to the MDS-UPDRS, we emphasize that our findings are applicable only to this scale.

Acknowledgments

Funding agencies: The Rush University Medical Center Section of Parkinson Disease and Movement Disorders receives funding and support from the Parkinson's Disease Foundation, New York, NY.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Engles JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol.* 2003; 56:968–976. [PubMed: 14568628]
- Molenberghs, G.; Kenward, MG. *Missing Data in Clinical Studies.* New York: Wiley; 2007.
- White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med.* 2005; 24:993–1007. [PubMed: 15570623]
- Fairclough DL, Cella DF. Functional assessment of cancer therapy (FACT-G): non-response to individual questions. *Qual Life Res.* 1996; 5:321–329. [PubMed: 8763800]
- Fairclough, DL. *Design and Analysis of Quality of Life Studies in Clinical Trials.* Boca Raton, FL: CRC Press; 2012.
- 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. [PubMed: 19025984]
- 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. [PubMed: 17115387]
- Goetz CG, Stebbins GT, Wang L, LaPelle NR, Luo S, Tilley BC. MDS-sponsored Scale Translation Program: process, format, and clinimetric testing plan for the MDS-UPDRS and UDysRS. *Mov Disord Clin Pract.* 2014; 1:97–101. [PubMed: 27747259]
- Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989; 45:255–268. [PubMed: 2720055]

10. Antonini A, Abbruzzese G, Ferini-Strambi L, et al. Validation of the Italian version of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale. *Neurol Sci.* 2013; 34:683–687. [PubMed: 22678179]
11. Martinez-Martin P, Rodriguez-Blazquez C, Alvarez-Sanchez M, et al. Expanded and independent validation of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). *J Neurol.* 2013; 260:228–236. [PubMed: 22865238]

Step 1: Review data set(s) to determine whether the missing data have been consistently or randomly dropped from the scale.

Step 2: Refer to Table 2 to determine the maximal number of permitted missed items for calculation of a final part score

Step 3: If the number of missing items exceeds the threshold, reject the rating and accept that the part total cannot be validly calculated with this technique. If the number of missing items is within the threshold, use the following formula to render an accurate part score (D):

$$(A \times B) \div C = D$$

A= Sum of item scores that have been recorded

B= Number of total items in the full part without missing values

C= Number of items with actual scores

D= Final calculated valid part score

Example 1: In a longitudinal population study, research personnel considered item 1.2 (Hallucinations and psychosis) as culturally insensitive, intrusive, and inappropriate for field workers to use in PD patients. This item was therefore dropped from the MDS-UPDRS part I on all cases. In a patient who had 12 of 13 Part I completed items but lacked an entry for item 1.2, the resulting part I score with missing value was 17.

What value should be placed as a total part I adapted score?

Solution: $(17 \times 13) \div 12 = 18$

Comment: With this technique, one item can safely be missing from a part I when that item is consistently missing in a data set. Data sets with more than one missing value in this instance, however would need to be discarded.

Example 2: In a clinical practice, a neurologist uses the MDS-UPDRS part III at all visits and keeps the total part score as an ongoing measure of motor severity. He works quickly, however, and finds that on five occasions over the past year in patient X, he has omitted items by mistake. On one busy day (4/22/14), he entered only 23 of 33 part III ratings, working quickly and forgetting to enter the scores after doing the tests. In reviewing the missing values, he considers that the dropped items were random on any given day due only to his fast pace of work.

For the entry on 1/21/14, 27 of 33 values are available (6 missing entries), and the sum of the available items is 42. What final part III score can he enter for this date?

Solution: $(42 \times 33) \div 27 = 51$.

Comment: A total part III score can be accurately rendered with seven or fewer randomly missing items.

For the entry on 4/22/14, he only entered 23 of 33 ratings for part III. The sum of those 23 ratings is 48. What final part III score can he enter for this date?

Solution: With this technique, no score can be generated because missing values exceed threshold.

Comment: The number of missing items is greater than seven, and this Part III score has to be rejected and left as "undetermined because of missing values."

FIG. 1.

Practical examples for clinical and research use derived from analyses presented.

TABLE 1

Number of allowable missing items to calculate MDS-UPDRS score when the same item is consistently missing across all patients

Part I: HY 1–2		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.963	0.979	0.979	0.988	
2	0.899	0.947	0.946	0.969	
Part I: HY 3		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.972	0.982	0.982	0.988	
2	0.916	0.949	0.948	0.964	
Part I: HY 4–5		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.967	0.978	0.978	0.985	
2	0.912	0.944	0.944	0.961	
Part II: HY 1–2		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.976	0.985	0.985	0.991	
2	0.943	0.967	0.966	0.980	
Part II: HY 3		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.972	0.982	0.982	0.987	
2	0.926	0.956	0.955	0.971	
Part II: HY 4–5		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.970	0.982	0.982	0.989	
2	0.924	0.945	0.945	0.965	
Part III: HY 1–2		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.990	0.994	0.994	0.997	
2	0.987	0.996	0.995	0.998	
3	0.978	0.993	0.992	0.997	
4	0.960	0.991	0.989	0.996	
5	0.942	0.988	0.986	0.995	
Part III-HY 3		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.992	0.995	0.995	0.997	
2	0.985	0.995	0.995	0.998	
3	0.967	0.993	0.992	0.997	
4	0.949	0.991	0.989	0.997	
Part III-HY 4–5		CCC			
No. of missing items	Min.	Median	Mean	Max.	

1	0.991	0.994	0.994	0.996
2	0.979	0.995	0.994	0.998
3	0.953	0.992	0.991	0.998
4	0.921	0.990	0.987	0.997
Part IV: HY 1–2		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.932	0.967	0.966	0.980
Part IV: HY 3		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.921	0.950	0.949	0.968
Part IV: HY 4–5		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.910	0.946	0.945	0.964

Legend: CCC for 480 PD patients (160 HY 1–2, 160 HY 3, and 160 HY4–5) when item scores from a given part of the MDS-UPDRS are consistently deleted as a missing value. White rows indicate that the minimal CCC falls at or above 0.95 and is acceptable. Gray rows indicate the point when the number of missing items renders a minimal CCC below threshold for valid calculation of the MDS-UPDRS total Part score. HY, Hoehn and Yahr Stage; CCC, Lin's Concordance.

TABLE 2

Number of allowable missing items to calculate MDS-UPDRS score when items are randomly missing across all patients

Part I: HY 1–2		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.981	0.989	0.989	0.994	
2	0.954	0.977	0.976	0.987	
3	0.933	0.962	0.961	0.980	
Part I: HY 3		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.979	0.988	0.988	0.992	
2	0.959	0.974	0.974	0.985	
3	0.932	0.958	0.957	0.974	
Part I: HY 4–5		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.976	0.984	0.984	0.990	
2	0.945	0.966	0.966	0.978	
Part II: HY 1–2		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.988	0.993	0.993	0.996	
2	0.970	0.984	0.984	0.991	
3	0.955	0.974	0.974	0.985	
4	0.934	0.962	0.961	0.977	
Part II: HY 3		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.982	0.989	0.989	0.993	
2	0.962	0.976	0.976	0.986	
3	0.934	0.961	0.960	0.975	
Part II: HY 4–5		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.984	0.990	0.990	0.994	
2	0.966	0.978	0.978	0.987	
3	0.944	0.964	0.964	0.978	
Part III: HY 1–2		CCC			
No. of missing items	Min.	Median	Mean	Max.	
1	0.996	0.998	0.998	0.999	
2	0.992	0.995	0.995	0.997	
3	0.987	0.992	0.992	0.995	
4	0.982	0.989	0.989	0.993	
5	0.977	0.986	0.986	0.991	
6	0.970	0.983	0.983	0.990	
7	0.967	0.979	0.979	0.987	

8	0.959	0.975	0.975	0.984
9	0.953	0.972	0.971	0.985
10	0.943	0.967	0.967	0.981
Part III: HY 3		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.995	0.997	0.997	0.998
2	0.991	0.995	0.995	0.997
3	0.987	0.992	0.992	0.996
4	0.980	0.989	0.988	0.993
5	0.974	0.985	0.985	0.992
6	0.971	0.982	0.981	0.990
7	0.961	0.978	0.978	0.986
8	0.955	0.974	0.974	0.983
9	0.951	0.970	0.969	0.982
10	0.941	0.965	0.965	0.981
Part III: HY 4–5		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.995	0.997	0.997	0.998
2	0.989	0.994	0.994	0.996
3	0.984	0.990	0.990	0.994
4	0.976	0.987	0.987	0.992
5	0.969	0.983	0.983	0.989
6	0.965	0.979	0.979	0.988
7	0.957	0.974	0.974	0.986
8	0.946	0.970	0.970	0.983
Part IV: HY 1–2		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.954	0.978	0.977	0.988
2	0.905	0.945	0.945	0.975
Part IV: HY 3		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.948	0.971	0.970	0.983
Part IV: HY 4–5		CCC		
No. of missing items	Min.	Median	Mean	Max.
1	0.948	0.970	0.969	0.982

CCC for 480 PD patients (160 HY 1–2, 160 HY 3, and 160 HY 4–5 when item scores from a given part of the MDS-UPDRS are randomly deleted as a missing value. White rows indicate that the minimal CCC falls above 0.950 and is acceptable. Gray rows indicate the point when the number of missing items renders a minimal CCC below threshold for valid calculation of the MDS-UPDRS total Part score. HY, Hoehn and Yahr Stage; CCC, Lin's Concordance.

TABLE 3

Maximal number of allowable missing items to calculate MDS-UPDRS total part scores

	Same Item(s) Consistently Missing Across All Patients	Different Item(s) Randomly Missing Across Patients
Part I: Across all HY	1	1
HY 1–2 only	1	2
HY 3 only	1	2
HY 4–5 only	1	1
Part II: Across all HY	1	2
HY 1–2 only	1	3
HY 3 only	1	2
HY 4–5 only	1	2
Part III: Across all HY	3	7
HY 1–2 only	4	9
HY 3 only	3	9
HY 4–5 only	3	7
Part IV: Across all HY	0	0
HY 1–2 only	0	1
HY 3 only	0	0
HY 4–5 only	0	0

The standard score for each patient's MDS-UPDRS (Part I, II, III, IV) score is used as the gold standard, and standard scores are calculated for each permutation with missing values. The threshold of a minimum CCC >0.95 is used to indicate the number of missing values allowable to provide a valid standard score even with missing data. HY, Hoehn and Yahr stages.

TABLE 4

Validation analysis for each MDS-UPDRS allowable number of missing values using patient data from official non-English versions of the MDS-UPDRS

Consistently Missing Items						
Part I	No. of missing items	HY Group	Min.	Median	Mean	Max.
	1	Stage 1–2	0.969	0.979	0.979	0.985
	1	Stage 3	0.975	0.982	0.982	0.987
	1	Stage 4–5	0.981	0.985	0.985	0.989
Part II						
	No. of missing items	HY Group	Min.	Median	Mean	Max.
	1	Stage 1–2	0.983	0.987	0.987	0.991
	1	Stage 3	0.984	0.989	0.989	0.991
	1	Stage 4–5	0.980	0.985	0.985	0.989
Part III						
	No. of missing items	HY Group	Min.	Median	Mean	Max.
	3	Stage 1–2	0.974	0.981	0.981	0.985
	3	Stage 3	0.968	0.976	0.976	0.982
	3	Stage 4–5	0.953	0.963	0.963	0.973
Randomly missing items						
Part I	No. of missing items	HY Group	Min.	Median	Mean	Max.
	1	Stage 1–2	0.984	0.988	0.988	0.992
	1	Stage 3	0.983	0.988	0.988	0.991
	1	Stage 4–5	0.986	0.989	0.989	0.992
Part II						
	No. of missing items	HY Group	Min.	Median	Mean	Max.
	2	Stage 1–2	0.978	0.985	0.985	0.990
	2	Stage 3	0.978	0.984	0.984	0.988
	2	Stage 4–5	0.973	0.982	0.982	0.986
Part III						
	No. of missing items	HY Group	Min.	Median	Mean	Max.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Consistently Missing Items				
7	Stage 1–2	0.976	0.984	0.983
	Stage 3	0.978	0.983	0.983
	Stage 4–5	0.972	0.981	0.981

From the total number of MDS-UPDRS scores from patients rated during the clinimetric validation program of official non-English versions of the MDS-UPDRS (N = 3,784), cases were divided by HY categories (Stages 1–2, 3, and 4–5). The smallest cohort was HY 4–5 with 380 entries, and, to balance groups, 380 from each category having more than 380 cases were randomly selected for this analysis (total sample, 1,140). The number of allowable missing items identified from the original cohort were tested against the non-English cohort to establish that the CCC in the non-English cohort for the missing items remained always above the threshold of 0.95. Data are shown for parts I–III confirming the exact fit. Likewise, as with the original analysis, all items' values in part IV must be present to generate an accurate score, because, even with 1 item score missing, the CCCs fall below threshold for consistently and randomly missing values (data not shown). HY, Hoehn and Yahr stage; CCC, Lin's Concordance.