RATER TRAINING ON UPDRS (PART III) MOTOR SECTION

Richa Gaur, PhD[1] , Hubert F Fernandez, MD[2], Ramon L Rodríguez, MD[2], Susan De Santi, PhD[3]

[1]The Cognition Group, Newark, United States, [2]Department of Neurology, University of Florida, Gainesville, United States,
[3]New York University School of Medicine, New York, United States

Abstract

Introduction
Training and evaluation of investigators, particularly on the motor section of the UPDRS, is critical before clinical trial initiation. This study examines an approach to rater training and certification on the UPDRS focusing on improving inter-rater reliability.

Method
283 raters from 34 countries participated in a rater training and certification program. This included a combination of online and live training at Investigator Meetings (IM). Web-based didactic training was followed by a practice session online. The IM included an intensive review of rating conventions and discussion of difficult to rate items. This was followed by a certification session using two patient videos for which raters provided their individual ratings to be compared with the gold consensus ratings (GCRs) set by two experts.

Results
Inter-rater reliability for Patient A ranged from .28 to .90, and from .27 to .94 for Patient B. Five raters for Patient A and ten for Patient B, had kappa levels which did not reach statistical significance. For Patient A, 23 items (85.1%) showed greater that 80% agreement and 15 items (55.6%) showed 90% agreement with the GCRs. For Patient B, 22 items (81.48%) showed greater than 80% agreement and 15 items (55.6%) showed 90% agreement.

Conclusions
The substantial inter-rater agreement of raters with the GCRs support uniform rating using the UPDRS scale as more than 80% of the items raters achieved significant rating accuracy. The results suggest that in a large PD trial intensive pre-study training on the UPDRS is effective in increasing the inter-rater reliability.

Introduction

The Unified Parkinson's Disease Rating Scale (UPDRS) is the most widely used scale for clinical trials in Parkinson's disease (PD).

The UPDRS includes: Part I – Mentation; behavior and mood; Part II – Activities of daily living (ADL); Part III – Motor examination (ME); Part IV – Complications of therapy.

Most clinical trials rely on the motor examinations in Section III as the primary outcome measure. Thus, rating accuracy of site raters on Section III, is critical to the study outcome.

A rater training and qualification program was conducted for two phase III multicenter studies on early and advanced PD.

The study examined an approach to rater training and certification on the UPDRS focusing on improving inter-rater reliability.

Methods

A global UPDRS (part III) rater training program included;

- online (web-based) training
- live training at the Investigator Meeting (IM).

The Raters

Training was provided before the start of the study to 229 raters from 199 sites representing 34 countries.

Rater's prior experience on UPDRS ranged from 1 month to 23 years (M = 6.2 years).

Training program

Training materials (subtitled into 19 languages) included five videos of English-speaking American patients (with early and advanced PD) interviewed by an American neurologist (figure 1).

Gold Consensus Ratings (GCRs) were set by two US based independent PD experts.

After completing the on-line training session ratings for Patient A were submitted online.

Data from raters' training session were presented at IM.

At the IM

Training by an expert trainer was provided on all sections of UPDRS, with emphasis on section III; and those items that were difficult to rate identified during the online training.

Certification at the IM

Raters observed and rated two videos; patient B (early PD) and patient C (advanced PD).

Raters had to meet the pass criteria for both videos in order to certify on the UPDRS.

Raters failing to certify received remediation with an expert trainer and additional certification attempt.

Results

Pass and fail status

Table 1 Pass and fail percentage of raters			
Patient	N	Pass N (%)	Fail N (%)
B	229	215 (94)	14 (6)
C	229	221 (96)	8(4)

Inter-rater reliability

Table 2 Inter Rater Reliability			
Patient	N	Kappa	P Level
B	229	.89	<.001
C	229	.64	<.001

For Patient B the statistically significant kappa levels for the individuals raters ranged from .28 to .78 with GCRs score (10 raters had kappa levels which did not reach the level of statistical significance).

For Patient C the kappa for the individual raters ranged from .27 to .94 with GCRs score (5 raters did not have significant kappa levels).

Table 3 Results of McNemar Tests for Binomial Proportion for items that were significantly different from the GCR.

	Test Items	% Matching Patient B	% Matching Patient C
1	**Speech**	67%	61%
2	**Facial Expression**	76%	87%
Tremor at Rest			
3	Face, Lips and Chin	84%	
4	Right Hand	80%	
6	Left Hand		78%
7	Right Foot	87%	87%
11	Left Foot	87%	
Action & Postural Tremor of Hands			
13	Right Hand	83%	
15	Left Hand	76%	
Rigidity			
16	Neck		75%
17	Right Upper Extremity	78%	
18	Left Upper Extremity	74%	
20	Right Lower Extremity	80%	
21	Left Lower Extremity		79%
Finger Taps			
22	Right	64%	
24	Left	80%	70%
Hand Movements			
25	Right		77%
27	Left		77%

Items with poor concordance between raters and GCR

Rater agreement with GCRs was analyzed by McNemar test for paired binomial proportions (table 3).

Previous experience on UPDRS and rater performance

Previous experience administering the UPDRS did not significantly affect scoring of patient C, or patient D.

Conclusions

Overall substantial to perfect agreement was obtained for patient B and C, suggests that raters rated the videos uniformly.

Speech, facial expression and tremor at rest (right foot) were items with least concordance with the GCRs.

Early PD patient was more difficult to accurately rate than an advanced PD patient as 13 items GCR for patient B and 9 items for patient C were significantly different than the GCRs.
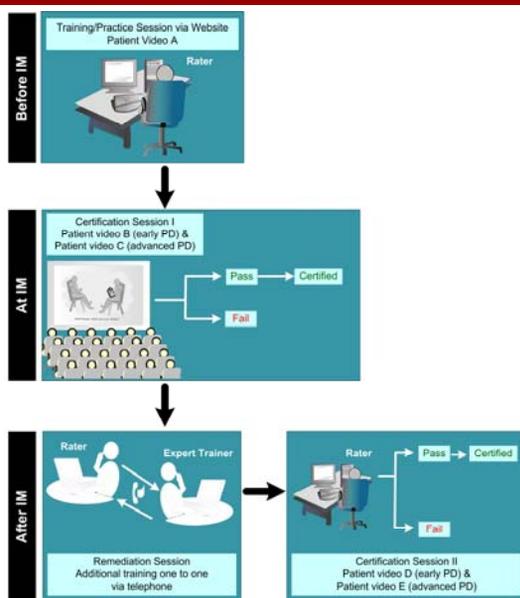
The difference in performance on two patient videos makes it critical for raters to undergo intensive rater training on UPDRS with emphasis on different patient vignettes.

Previous experience administering the UPDRS did not significantly affect scoring of patient C, or patient D.

References

1. Goetz, C.G (2003). The Unified Parkinson's Disease Rating Scale (UPDRS): Status and Recommendations. *Movement Disorders* Vol. 18, No 7 pp 738 -750.

2. European Medicines Agency (2008). Guideline on clinical investigations of medicinal products in the treatment of Parkinson's disease. London UK .

Figure 1 Training and certification process