Does Rater Experience Influence UPDRS Inter-rater Reliability?

Alan Kott, MD¹; Jina Swartz, MD, PhD²

¹Bracket, Prague, Czech Republic; ²Bracket, Goring-on-Thames, United Kingdom
Background: Clinical trials using subjective outcome measures such as the UPDRS in Parkinson’s disease (PD) clinical trials (Fahn et al., 1987) are dependent on the validity and reliability of collected data. When initiating a clinical trial, despite calls for proper inter-rater reliability assessments, a surrogate measure such as experience level is often used as the only criterion to allow raters to assess subjects in the study.

There is discrepant information on the influence of rater experience on the UPDRS inter-rater reliability. We have therefore investigated the effect of various experience indices on the levels of inter-rater reliability.

Methods: Any rater intending to rate study subjects in either of the two studies was required to undergo a training and certification exercise and be certified accordingly.

- Rater training consisted of didactic and interactive training on each of the UPDRS part III items utilizing PowerPoint slides and multiple video examples or vignettes on each item.
- Rater certification consisted of a scoring of a videotaped full UPDRS part II assessment, excluding the rigidity item. Each study had a unique video developed.

Results: The overall kappa results by study are presented in Table 1. Tables 2 and 3 summarize kappa results broken down by Degree and Study (Table 2 and Graph 1) and Frequency of UPDRS Usage and Study (Table 3 and Graph 2). In Study #1 no significant correlations were identified between kappa results and years of clinical experience (r(284) = 0.0154, p = 0.7395) or between kappa values and years of clinical trial experience (r(283) = 0.0160, p = 0.7884). Similarly non-significant results were obtained in Study #2 for correlations between kappa results and years of clinical experience (r(367) = 0.0004, p = 0.9219) as well as between kappa values and years of clinical trial experience (r(366) = 0.0198, p = 0.7199) (Graphs 3 and 4).

Conclusions: We have conducted a retrospective analysis of certification data collected for two phase III Parkinson’s disease studies, aiming to identify whether there is any relationship between experience level indices and inter-rater reliability. Our analyses indicate that experience has very limited, if any, effect on inter-rater reliability and should therefore not be used as a surrogate means of evaluation when selecting raters. Our data indicate that significant inter-individual variability of rater kappa values exists even between the most experienced raters. It is therefore important to assess the inter-rater reliability at the beginning of each clinical trial and select only those raters who have achieved good kappa values. Due to potential decrease in reliability during the study quality assurance methods, such as blinded data analysis on the individual rater and site levels, need to be employed.

References:

Disclosures: The poster is financially supported by Bracket (formerly Specialty Clinical Services of United BioSource Corporation). Both authors are employed by Bracket.

Graphs 1-4: Kappa Values by Degree (MD vs Non-MD) and Study, Kappa Values by UPDRS Usage and Study, Kappa Values by Experience Level and Study, and Kappa Values by Years of Clinical Experience.