Inter-rater Reliability and Error Analysis of the Scales for Outcomes of Parkinson's Disease - Cognition (SCOPA-COG) in MODERATO: A randomized double blind placebo controlled study to assess the effect of rasagiline on mild cognitive impairment in PD

Robin C Hilsabeck[1], Kari R Nations[2], Russell Tanenbaum[1], Beth Grubb[3], Azhar Choudhry[4],

[1]INC Research, Austin, Texas, United States; [2]INC Research, Austin, Texas, United Kingdom; [3]Teva Branded Pharmaceutical Products R&D, Inc., Horsham, Pennsylvania, United States; [4]Teva Branded Pharmaceutical Products R&D, Inc, Horsham, Pennsylvania, United States.

Background: The SCOPA-COG was developed for use in clinical research and has demonstrated good internal consistency, test-retest reliability, and construct validity. However, inter-rater reliability (IRR) of the SCOPA-COG is unknown. The purpose of the present study was to examine IRRs and provide detailed analyses of administration and scoring errors. **Methods:** The MODERATO study is a multi-center RCT examining the effect of rasagiline on cognition in patients with PD. The SCOPA-COG, which consists of 10 items that comprise four subscales, is the primary endpoint. In an effort to maximize IRR, a robust training program was employed consisting of on-line didactic and video training and videotaped administrations to mock patients that were reviewed by calibrated expert raters who provided corrective feedback. Following successful completion of SCOPA-COG training, raters were "certified" to begin testing study subjects. Each study subject's video was reviewed by an expert rater who provided feedback and was blinded to the site rater's scores. Thus, IRRs based on expert and site rater scores were calculated for the total score, subscales, and individual items. In addition, IRRs for site raters who had seen multiple subjects were computed to explore the impact of experience and corrective feedback over time. Finally, detailed analyses of the types of errors affecting subjects' scores were conducted. **Results:** At submission, 25 site raters had conducted videotaped SCOPA-COG administrations on 88 subjects, which had been reviewed by two expert raters. Nineteen of the site raters had performed at least two SCOPA-COG subject administrations. The overall IRR for the SCOPA-COG total score was .98, and IRRs for individual items ranged from .85 (Fist-Edge-Palm) to 1.0 (Delayed Recall). Errors affecting subjects' scores occurred in 48% of cases, with the most errors occurring on Fist-Edge-Palm, followed by Indicate Cubes. Over time, IRRs increased and errors decreased. **Conclusions:** Results indicate that intensive rater training on the SCOPA-COG can produce good to excellent IRRs. However, almost half of the cases contained errors that affected subjects' scores. Findings suggest that continuous training and surveillance on the SCOPA-COG is valuable, and emphasis can be placed on particular items to minimize error and increase study power.