Performance During "Applied" Training on the MADRS/SIGMA as a Predictor of Change in a Global Depression Trial

Rothman, B[1], Lord-Bessen, J[1], Sanchez, R[2], Peters-Strickland, T[2], Opler M[1,3]

[1]ProPhase LLC, New York, NY

[2]Otsuka Pharmaceutical Commercialization and Development, Princeton, NJ

[3]New York University, New York, NY

Methodological Question: This analysis was conducted to identify which aspects of applied training performance were most predictive of MADRS change over time in a global depression trial.

Introduction/Aims: Many rater training programs rely on passive scoring tasks, such as video interview scoring, to assess inter-rater reliability. It is widely accepted, however, that monitoring and review of a rater's "applied" performance (e.g., via a simulated interview) is a more accurate and ecologically valid way to judge rater quality and thus his or her ability to detect change over time. When accompanied by standardized tools (e.g., the Rater Applied Performance Scale), MADRS evaluators can assess a rater's capacity to administer and score an assessment in a way that better approximates in-study behavior. Using data from a multi-site, global depression trial, we sought to identify which domains of applied performance were most predictive of change in the MADRS.

Methods: Eighty qualified raters participating in a global depression trial completed a 1-1 "applied" training session via telephone as part of pre-study certification on the MADRS. Each interview used the SIGMA and was scored according to six RAPS domains (Lipsitz, et al., 2004) – Adherence, Follow-up, Clarification, Neutrality, Rapport, and Accuracy of Scoring. Results from applied training were compared to in-study data from 191 subjects, assigned to one of two active comparator arms (escitalopram + placebo or aripiprazole + placebo) or to a combination therapy arm (aripiprazole + escitalopram).

Results: Of the six RAPS domains, both Follow-Up (r=-0.17, p=0.0031) and Scoring (r=-0.29, p=0.000) were negatively correlated with absolute value of change in MADRS. For all domains but one, the higher the score on the domain, the less change in MADRS from Week 8 to 14. The only domain with a positive relationship with change was Rapport, but it was not significant (r=.06, p=.489).

Discussion/Conclusions: In this sample, Follow-Up and Accuracy were the only RAPS domains significantly correlated with change in MADRS scores. This suggests that fidelity to the structured interview guide (adherence), as well as to scoring conventions, are the most robust predictors of change and thus a rater's ability to detect signal. Training and in-study surveillance programs involving the MADRS should focus on these domains when evaluating rater quality.

Disclosure: This project was funded by Otsuka Pharmaceutical Commercialization and Development and by ProPhase, LLC. The authors report no conflicts of interest for this work.