PW01-140 - **THREE LEVELS OF RATER PERFORMANCE IN STANDARDIZED PANSS TRAINING**

G. Lo[1], **W. Yavorsky**[2], M. Opler[3], A. Defries[4]

[1]*ProPhase LLC & Fordham University,* [2]*ProPhase LLC,* [3]*ProPhase LLC & New York University,* [4]*ProPhase LLC & Teachers College, Columbia University, New York, NY, USA*

Objectives: The importance of rater training for reliability and validity in clinical trials is well documented. In this study the authors examined data from several large training events that used standardized training procedures and raters with similar levels of education and experience to determine if there were any differences between rater performance at baseline (before training) and endpoint (after training) that emerge independent of these factors.

Methods: Results from multiple training events held internationally were analyzed to determine if differences between baseline and endpoint scores were significant. Raters scored videotaped interviews of the Positive and Negative Syndrome Scale (PANSS).

Results: Three subgroups of raters emerge based on concordance with gold-standard ratings: raters that score high at baseline and endpoint; raters that score fair at baseline and good at endpoint; raters that do not appear to improve. For the stronger initial raters that continued to perform well after training there was a less substantial change ($t(96)=2.953$, $p< .005$) than those raters that did less well at baseline but improved after training ($t(160)=4.037$, $p< .001$).

Conclusion: Within training groups there appear to be three groups of raters that emerge independently of rater qualification and training received: those that performed well initially and well at endpoint; another group that performed marginally at the beginning of training and showed improvement by the end; and a third group that did not improve as a result of training. Analysis of ICCs suggests targeted training for individuals that perform less well at baseline could be beneficial.