Introduction

- Increasingly, academic and industry researchers have been using scale specific algorithms – pre-set rules of item association – to evaluate psychometric study data for risks associated with rater error. Rater error is a significant source of variance (Alphs et al, 2012) in clinical trials and the data-monitoring process is designed to detect problematic scale use and provide correction during the course of a trial. These rules were developed using the factor structures of the instrument (e.g., Marder et al, 1997) wherein, for example, items that measure aspects of given factor should be scored in a particular manner given directionality and magnitude of comparable items. If the direction and magnitude of two items are sufficiently inconsistent, it is referred to as a “violation.” If such a violation occurs then the data surveillance organization contacts the clinician to determine if the violation was a true positive and an error had been committed. The results of these contacts has been collected over a six year period and categorized into binaries of true and false positives. The purpose of this study was to use ROC analysis to evaluate the sensitivity of such algorithms to detect the risk of problematic administration of the PANSS in clinical trials in light of the empirical information collected within study.

Objectives

- To determine the sensitivity of data monitoring algorithms to detect error using empirical outcome evidence about actual rater error from a pooled data set.

Methods

- ROC analysis (Florkowski, 2008) was used to determine the true-positive rate, or how sensitive the data-monitoring algorithms are in the detection of problematic data where it existed. The monitoring algorithms split the sample into binary data that contained item relationships that violated the scale logic (indicative of rater error) and those that did not. Results of all contacts with clinicians were housed in a proprietary data capture system conforming to 21 CFR Part 11. R 3.1.1 statistical software was used to conduct the ROC analysis and data was considered against the background of dichotomized empirical results from the trials, i.e., if the presence of violation was verified through contact with the rater.

Results

- Pooled data (n=4096) from eight Phase III schizophrenia trials that had been subject to in-study data-monitoring was used for this analysis. The true positive rate was found to be in the very good range GLM (AUC = .81) given the model parameters RF (AUC =1). In figure 1 the dotted line represents the reference line or, what would be expected by chance. The green line is what would be expected at perfect classification. The blue line is the true positive rate indicating that the algorithms predicted rater error when it actually existed based on contact with the rater.

Conclusion

- These results suggest that algorithms designed to detect the error in the PANSS, in their current characterization, appear to have strong predictive power to determine risk of rater error in clinical trials for schizophrenia. The results of improving rater accuracy in-study can lead to better differentiation between drug and placebo and serve to mitigate the risk of both Type-I and Type-II error.

References