Introduction

- Increasingly, academic and industry researchers have been using algorithms – pre-set rules of association – to evaluate study data for risks associated with rater error. Though several organizations have employed this type of system for many years, there has been a special working group set up to investigate these algorithms as they apply to the PANSS. The assumption is that by looking at expected relationships between scale items as well as the clinical likelihood of change across visits for a given indication, mistaken or idiosyncratic use of scale can be detected and ameliorated during the course of the study.

- The purpose of this study was to use ROC analysis to evaluate the sensitivity of algorithms to detect the risk of problematic administration of the PANSS in clinical trials. We did not feel that it was problematic to pool this data as the variable under consideration was not related to the type of patient or compound but rather discrepancies in data as proxy for rater performance. There were four primary categories of algorithm: intra-item discrepancies; across scale discrepancies; across visit inconsistencies and clinically improbable change and/or presentation.

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

- Data monitoring was conducted for the duration of each of the eight Phase III clinical trials. ROC analysis was used to determine the true-positive rate, or how sensitive the data-monitoring algorithms are in the detection of problematic data.

- The monitoring algorithms split the sample into binary data: either the item relationships reflected violations of the categories identified (indicative of rater error) or they did not. 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.

- 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.

Results

- 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.

Conclusion

- Increasingly, academic and industry researchers have been using algorithms – pre-set rules of association – to evaluate study data for risks associated with rater error. Though several organizations have employed this type of system for many years, there has been a special working group set up to investigate these algorithms as they apply to the PANSS. The assumption is that by looking at expected relationships between scale items as well as the clinical likelihood of change across visits for a given indication, mistaken or idiosyncratic use of scale can be detected and ameliorated during the course of the study.

- The purpose of this study was to use ROC analysis to evaluate the sensitivity of algorithms to detect the risk of problematic administration of the PANSS in clinical trials. We did not feel that it was problematic to pool this data as the variable under consideration was not related to the type of patient or compound but rather discrepancies in data as proxy for rater performance. There were four primary categories of algorithm: intra-item discrepancies; across scale discrepancies; across visit inconsistencies and clinically improbable change and/or presentation.

- 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

- Data monitoring was conducted for the duration of each of the eight Phase III clinical trials. ROC analysis was used to determine the true-positive rate, or how sensitive the data-monitoring algorithms are in the detection of problematic data.

- The monitoring algorithms split the sample into binary data: either the item relationships reflected violations of the categories identified (indicative of rater error) or they did not. 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.

- 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.