Impact of Centralized Over-Read on Outcomes in Depression and Schizophrenia Trials

Barbara Echevarria PhD[1], Selam Negash PhD[1], Michael T. Ropacki PhD[1]
[1] MedAvante-ProPhase, Inc.

THE METHODOLOGICAL QUESTION BEING ADDRESSED

Centralized over-read (Central Review) of clinical assessments in depression and schizophrenia trials implemented at key visits (e.g., screening, baseline and end of study) is intended to improve interrater reliability. However, the extent to which this approach reduces scoring variability in outcome measures at key visits has not been adequately explored. Therefore, we investigated the impact of performing Central Review on two widely used efficacy measures administered in schizophrenia and depression trials, the Positive and Negative Syndrome Scale (PANSS) and Hamilton Depression Rating Scale (HAM-D), respectively.

INTRODUCTION (AIMS)

- Scoring variability and poor interrater reliability has been reported to contribute to high placebo response rates and inconclusive results in clinical trials[1,2] and rater training has been demonstrated to be an insufficient remedy to these challenges[3].
- Among key study visits, scoring variability is a concern: at screening, which determines whether a subject is to be included in the study based on a certain scale score; at the baseline visit, which sets the score against which treatment efficacy will be measured; and end of study, which establishes whether the treatment provided to the subject was efficacious.
- Quality assurance measures such as Central Review methodology have been implemented to avoid rater drift and improve inter-rater reliability, but the impact of this approach on scoring variability and rater agreement in PANSS and HAM-D assessments has not been adequately explored.
- The present study investigated the impact of Central Review on total scale scores for key study visits (screening, baseline and end of study) in two multisite trials of depression and schizophrenia using the HAM-D and PANSS.

METHODS

- Data from PANSS and HAM-D assessments in two separate randomized, double-blind, multisite schizophrenia and depression clinical trials were analyzed.
- The scales were completed by site raters rigorously trained on administration and scoring conventions who were then qualified by successfully completing a rating precision exercise prior to conducting in-study assessments.
- A cohort of expert calibrated Central Review clinicians examined video/audio recordings and source documents to identify raters' administration and scoring errors.
- When scoring discrepancies were identified, raters were given an option to either agree with the reviewer's feedback and change the score, or provide rationale for rejecting the feedback and maintain the original score.
- The number of scoring discrepancies and score changes that resulted from Central Review feedback for screening, baseline and end of study visits in both trials was examined.
- The impact of score changes on PANSS and HAM-D total scale score for key study visits was also analyzed.

RESULTS

- A total of 2,057 PANSS and 358 HAM-D assessments were reviewed. The mean and standard error (SE) for each overall sample were 94.3 (13.3) for the PANSS and 17.9 (6.5) for the HAM-D total scale scores.
- Error reduction as a function of central review was examined for both measures. Figure 1 shows that scoring discrepancies declined from screening (27%) to end of study visits (16%) for the PANSS $t(839) = 2.9$, $p = .004$.
- There was also a trend towards a decline in scoring discrepancies for HAM-D (40% to 23%, screening to week 8, respectively) that did not reach significance, $t(141) = 1.6$, $p = 0.1$.
- Absolute score changes in the total scores are shown in Figure 2. Central Review resulted in a mean total score change on average of three points for the PANSS and two points for the HAM-D at key study visits.
- Most score changes in the PANSS were to higher scores after Central Review, regardless of the visit. For the HAM-D, site raters tended to score higher than the Central Review clinicians for screening and baseline visits, and lower at the end of study visit.

Figure 1. Percentages of Reviews with Discrepancies Across Key Study Visits

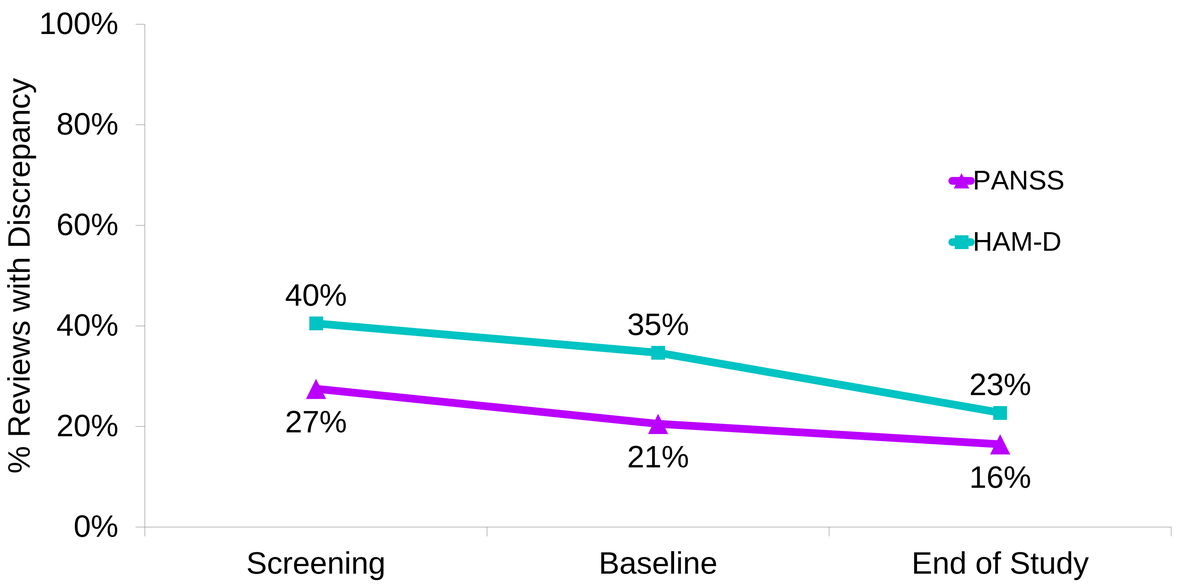
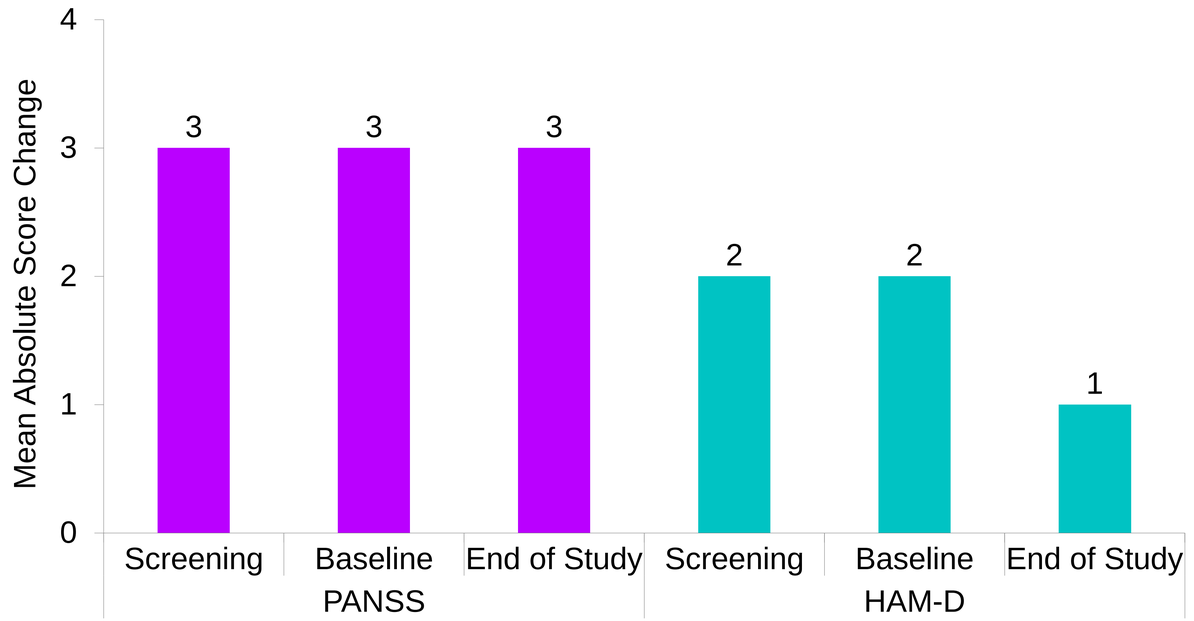


Figure 2. Mean Absolute Score Changes as a Function of Central Review



CONCLUSIONS

- Supplementing initial rater training with additional in-study oversight by expert and highly calibrated Central Review clinicians resulted in a positive impact on individual item scores and total scale score for key study visits, improving interrater reliability and reducing rater drift in schizophrenia and depression trials.
- Central Review of screening visits can improve subject selection.
- When applied to baseline and endpoint visits, Central Review can enhance data reliability. Additionally, Central Review feedback improves reliability in scoring of frequently used endpoint measurements in psychiatry trials.
- The reduction of scoring discrepancies over time proves that Central Review is a valid method of providing in-study training that extends the effectiveness of initial rater training, which is not sufficient to ensure rating reliability during the entire life of the study. Despite continued feedback throughout the study, monitoring of key late-study visits such as the end of study visit, is still recommended, as scoring errors can occur at any time.
- Reduced error variance translates to increased study power and a higher probability of signal detection in interventional trials.

References:
1. Khan A, Yavorsky WC, Liechti S, DiClemente G, Rothman B. Assessing the sources of unreliability (rater, subject, time-point) in a failed clinical trial using items of the Positive and Negative Syndrome Scale (PANSS). *J Clin Psychopharmacol*. 2013 Feb; 33(1):109-17.
2. Kobak KA, Kane JM, Thase ME, Nierenberg AA, 2007. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol*. Feb;27(1):1-5.
3. Targum SD, Evaluating Rater Competency for CNS Clinical Trials. *J Clin Psychopharmacol*. 2006; 26:308-310.

MedAvante-ProPhase

A WIRB-Copernicus Group Company