The Importance of Rigor in Post-baseline Assessments in CNS Trials

Williams, JBW[1,2] Popp, D[1] Kobak, KA[1,3] Detke, M[1,4]

MedAvante, Inc.[1], College of Physicians and Surgeons, Columbia University[2], Center for Psychological Consultation[3], Indiana University School of Medicine[4]

INTRODUCTION

Inappropriate subject selection, "functional unblinding," rater drift, and expectation biases can all contribute to trial failure. At baseline, pressure to enroll subjects can cause either inflation of baseline severity scores or inappropriate diagnoses. An increasing number of studies now include some method, such as the use of independent blinded raters, for ensuring that the right subjects are entered into a trial.

Post-baseline factors, however, can also affect trial outcomes. Expectation biases, functional unblinding, and rater drift can obscure a drug-placebo difference or introduce Type II (false negative) errors. Familiarity with a subject over the course of a study can influence a rater's scoring and create expectation biases. Expectation bias can increase with the amount of time the subject spends at the trial site[1]. Observing adverse events may lead one to assume the subject is on the investigational drug or active comparator (functional unblinding). Measurement noise or variability is introduced by drift from standardized scoring conventions among raters over a study's duration. Independent raters, blinded to the study treatment and study visit (i.e., screening, baseline, or endpoint) and continuously calibrated may mitigate expectation biases, functional unblinding, and rater drift.

KEY FOR RATING SCALES

PANSS = Positive and Negative Syndrome Scale
BPRS = Brief Psychiatric Rating Scale
MMSE = Mini-mental State Examination
SAPS = Scale for the Assessment of Positive Symptoms
NPI = Neuropsychiatric Inventory

METHODS

Studies with ratings by both site raters and blinded independent central raters can be evaluated to see how critical are continued blinding and continuous calibration after baseline visit. In a trial of acute schizophrenia, blinded independent central raters conducted PANSS interviews and site raters applied the BPRS on the same 313 subjects. A study of 287 subjects with Parkinson's psychosis included blinded independent central ratings in the US, and traditional site ratings in the sites outside of the US (OUS). Both sets of raters used subscales of the SAPS as the primary outcome measure. A negative trial of Generalized Anxiety Disorder (GAD) had blinded independent raters evaluate 122 subjects assigned to the placebo arm who had been admitted to the study by site raters' SIGH-A baseline evaluations.

Study #1: Acute Schizophrenia[2]

Sample: N=313 hospitalized subjects with acute schizophrenia
Study inclusion: PANSS ≥ 70 and ≤ 120 by blinded centralized raters
Primary outcome measure: PANSS by blinded independent central raters (converted to a derived BPRS score)
Other measures: BPRS by site raters, with access to blinded raters' PANSS scores (blinded raters always went first)
Study duration: Six weeks of treatment
Study arms: Placebo, active comparator (olanzapine), and two experimental drug treatment arms (off-label)
Blinded independent continuously-calibrated raters: N=18
Traditional site raters: 35 sites

Study #2: Parkinson's Psychosis[3]

Sample: N=287 subjects with Parkinson's psychosis
Study inclusion: MMSE of <21; symptoms severe enough to warrant treatment with an antipsychotic agent as documented by items A and B of the NPI, and defined as the sum of Hallucinations (frequency × severity) and Delusions (frequency × severity) ≥ a total score of 4
Primary outcome measure: SAPS subscales for hallucinations and delusions
Study duration: Six weeks of treatment
Study arms: Placebo, 10 mg pimavanserin (off-label), 40 mg pimavanserin (off-label)
Blinded independent continuously-calibrated raters in US: N=11; 50 sites in the US (no site raters in US)
Traditional site raters: 36 sites outside of the US (OUS) (no blinded raters OUS)

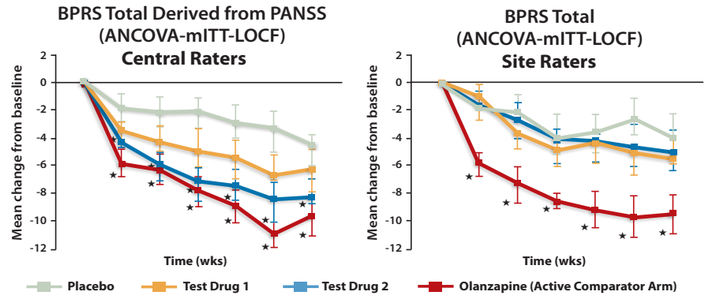
Study #3 : Generalized Anxiety Disorder[4]

Sample: N=122 subjects with Generalized Anxiety Disorder in placebo arm
Study inclusion: HAMA ≥20 at screen and baseline and ≥2 on HAMA items 1 and 2 by site raters
Primary outcome measure: SIGH-A by site raters
Other measures: Blinded independent central raters interviewed at baseline and at week six; at baseline site raters always went first; at week six, rater order was counterbalanced
Study duration: Six weeks of treatment
Study arms: Placebo, 0.9 mg/day experimental dose (off-label), and 1.5mg/day experimental dose (off-label)
Blinded independent continuously-calibrated raters: N=22
Traditional site raters: 119 raters at 45 sites

RESULTS

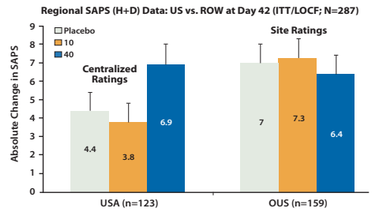
Study #1 : Acute Schizophrenia

- Blinded independent central raters observed statistically significant efficacy at every time point after week one in one experimental dose arm.
- Site ratings failed to show efficacy in either dose arm at any time point.



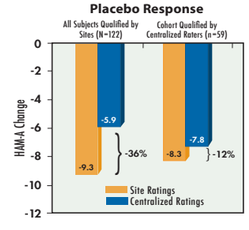
Study #2 : Parkinson's Psychosis

- Blinded, independent central raters in the US observed statistically significant efficacy at two weeks and a trend toward significance at six weeks in the 40 mg dose arm.
- Site ratings outside the US did not separate any drug dose from placebo.



Study #3 : Generalized Anxiety Disorder

- Active treatments failed to separate by any measure; no positive control included.
- Analysis focused on placebo response.
- In overall population (N=122) placebo response was 36% lower as assessed by blinded independent central raters. Also, in the cohort included by the central raters (n=59) placebo response was 12% lower as assessed by centralized ratings.



CONCLUSIONS

The Schizophrenia study demonstrates that post-baseline ratings performed by blinded independent central raters did detect separation of drug and placebo in study arms where site ratings did not. Since both central ratings and site post-baseline ratings were conducted with the same subjects, the outcome improvements may be attributed to differences in post-baseline assessments. In the Parkinson's psychosis study, blinded independent central raters found differences between drug and placebo while OUS site raters did not observe any separation. Finally, in the GAD study, blinded independent central raters observed smaller placebo response than raters based at the sites. We hypothesize that familiarity with the subject, observation of adverse events, and knowledge of visit sequence may have led to expectation biases on the part of site raters regarding degree of change or in which treatment arm a subject was enrolled.

Data from several studies now support the importance of the accuracy of outcome assessments after subject selection, even when subject selection is performed by centralized raters. Continued vigilance and precision of ratings beyond baseline can increase the sensitivity of findings in a clinical trial and decrease placebo response rates. Blinding of raters to study protocol and visit number, and independence from subjects' sites minimizes expectation bias; non-specific treatment effects are reduced by limiting the rating staff interactions with subjects. Rater drift, even with experienced raters, can be diminished only through continuous calibration of the cohort of raters.

References:

1. Guico-Pabia CJ, Musgnung J, Pederson R, Ninan PT: Placebo response in trials of antidepressants in patients with major depressive disorder. Poster presented at the annual conference of the National Clinical Drug Evaluation Unit 2010.
2. Data on file MedAvante, Inc.
3. Friedman JA, Ravina B, Mills R, Williams H, Bahr D, Peters P, Tison F, Burn D: Pimavanserin phase III PDP results. Presented at the annual conference of the American Academy of Neurology 2010.
4. Williams JBW, Dunn J, Kobak KA, Giller E, Curry L, Wilson P, Detke M: Placebo response assessed by site and remote blinded centralized raters in a GAD trial.

Disclosures:

Williams, JBW, MedAvante Inc. Kobak, A, MedAvante Inc., Center for Psychological Consultation. Detke, M, MedAvante Inc., NIH, Denysias, Inc., Sonkei Inc., Insight Neuropharma, Inc., Jeevan Scientific, Inc., Pam Lab, Inc., Columbia NW Pharmaceuticals, Naurex, Inc.

MEDAVANTE® RESEARCH INSTITUTE
Calibrated for Certainty