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

INTRODUCTION

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 make expectation biases more evident. Expectation bias can increase with the amount of time the subject spends at the trial site. Observing adverse events may lead one to assume the subject is on a drug/placebo that is different from what they are actually on. Inappropriate subject selection, “functional unblinding,” rater drift, and measurement noise or variability is introduced by drift from standardized scoring conventions among raters over a study’s duration. Independent blinded raters, blinded to the study treatment and study visit (i.e., screen- ed raters), blinded to 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 blinded raters, blinded to the study treatment and study visit (i.e., screen- ed raters), blinded to the investigational drug or active comparator (functional unblinding).

METHODS

Studies with ratings by both site raters and blinded independent central raters can be evaluated to see how critical are continued blinded 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 raters in 35 sites.

RESULTS

Study #1: Acute Schizophrenia

- Sample: N=313 hospitalized subjects with acute schizophrenia
- Study inclusion: PANSS ≥ 70 and ≤ 120 by blinded central raters
- Primary outcome measure: PANSS by blinded independent central raters (converted to a derived BPRS score)

Study #2: Parkinson’s Psychosis

- Sample: N=287 subjects with Parkinson’s psychosis
- Study inclusion: MMSE ≥ 22; 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: n. severity; and delusions (frequency: n. severity) ≥ a total score of 4
- Primary outcome measure: SAPS subscale for hallucinations and delusions
- Study duration: Six weeks of treatment
- Study arms: Placebo, 40 mg pimavanserin (off-label), 40 mg pimavanserin (off-label)

Study #3: Generalized Anxiety Disorder

- Sample: N=122 subjects with Generalized Anxiety Disorder in placebo arm
- Study inclusion: NIMH-21 at screen and baseline and ≥2 on HAM-A 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 first, at week six, rater order was counterbalanced

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 only when site raters did not observe any separation. Finally, in the GAD study, blinded independent central raters observed smaller placebo response compared to site raters when the data were blinded. We hypothesize that familiarity with the subject, observation of adverse events, and knowledge of trial sequence may have led to expectation biases on the part of site raters regarding degree of change in which treatment arm a subject was enrolled.

Data from several studies now support the importance of the accuracy of outcome assessments after subject baseline. 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 minimize exposure 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


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 retention 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.