Title: Variation in Diagnostic Reliability Across MDD Clinical Trials

Authors: Sachs GS[1,2], Reksoprodjo, P[1], Roy, M[1], Kott, A[1], Busner J[1,]

Affiliations: [1]Bracket. [2.] Massachusetts General Hospital

Methodological Question: Does diagnostic reliability as measured by agreement between categorical diagnostic instruments and continuous severity rating scales impact response to treatment?

Aims: Demonstrate that a diagnostic criteria and a threshold score on the Montgomery–Åsberg Depression Rating Scale (MADRS) produce congruent results at study entry. We expected that subjects with a reliable diagnosis of Major Depressive Episode (MDE) have at least four MADRS items scored ≥ 4 and at least one of the mood items or the anhedonia item rated at or above the threshold for clinical significance.

Method: Blinded data from 9 large double-blind, randomized Major Depressive Disorder studies requiring a current MDE at study entry were analyzed. MADRS items were considered clinically significant if scored ≥ 4. Gating criteria were considered to be insufficient if none of the MADRS items corresponding to DSM gating criteria was rated ≥ 4. Insufficient psychopathology for MDE was defined as having fewer than four DSM criteria rated clinically significant based on the corresponding MADRS items. Unreliable diagnosis was defined as meeting DSM criteria by a diagnostic instrument but lacking gating criteria or insufficient psychopathology for a MDE.

Results: Overall 3,256 subjects were identified with screening and baseline assessments. Nearly all (n=3,091 = 95%) were randomized. Among those randomized, 585 (18.9%) met our operational criteria for unreliable diagnosis, 91 (2.9%) lacked gating criteria, and 579 (18.7%) had insufficient psychopathology for MDE. The proportion of unreliable diagnosis varied widely across studies (7.3% - 37.4% ($\chi^2(8) = 164$; $p < 0.001$).

Blinded data for MADRS change from baseline was available for five studies. In these studies subjects with reliable diagnosis had significantly higher baseline MADRS scores and had significantly greater MADRS change scores (15.7) vs those classified as having an unreliable MDE diagnosis (12.5).

Conclusion:

This blinded data analysis indicates that a substantial proportion of subjects entering studies may have unreliable diagnosis. All subjects were considered to meet MDE criteria at screening and baseline according to diagnostic assessment used for subject qualification. These results are, however, not concordant with expectations from continuous measures of symptom severity.

The variation across studies in the rate of randomizing subjects with unreliable diagnosis suggest the discordance between categorical assessment of symptom criteria and continuous ratings of severity are more likely to be associated with variance in quality metrics rather than an artefact of the shortcomings in the correspondence between MADRS items and the DSM MDE criteria and supports the use of diagnostic validation procedures in clinical trials.

The difference in the blinded MADRS change scores is of a magnitude (Δ 3.2 points) similar to the effect size used to calculate sample size requirements in these studies, but may be confounded by the baseline differences between subjects with reliable vs unreliable MDE diagnosis.

Further analysis with unblinded data are needed to compare signal detection in subjects with less reliable diagnosis to subjects with high levels of diagnostic confidence.

Disclosures: All authors are full time employees of Bracket