Independent and Site-based Ratings of Symptom Severity in Pharmacogenetics Clinical Trials

Selam Negash[1], Bryan Dechairo[2], Lindsey Burns[2], Briana Webber-Lind[1], Krystal Brown[2], Jennifer Lord-Bessen[1], Mark Opler[1,3]
[1]MedAvante-ProPhase, Inc. [2]Myriad Genetics [3]NYU Medical Center

BACKGROUND

- There has recently been a resurgence in innovation in psychiatry, from the development of new mechanisms of action to the use of technologies to evaluate and monitor treatment[1].
- Alternatives to standard efficacy evaluations are required to reduce the impact of treatment unblinding and alleviate the effect of therapeutic expectation and other phenomena that may reduce signal detection.
- One area of particular promise is that of personalized medicine and pharmacogenomics. Studies in this area require novel approaches to ensure endpoint reliability and validity, reduce the impact of treatment unblinding, and minimize other confounds that reduce signal detection.
- The current program utilized a combination of novel approaches to ensure endpoint fidelity, including the use of remote, independent ratings to evaluate adherence to inclusion criteria and efficacy in a trial of major depressive disorder[2].

METHODS

- Analyses were carried on two endpoints: the 16-item Quick Inventory of Depression Symptomology (QIDS-C16)[3] and Hamilton Depression Rating Scale-17 (HAMD-17)[4]. The QIDS-C16 was used at screening, while both scales were administered at baseline and follow up visits. Baseline QIDS-C16 scores ≥11 were used as inclusion cutoffs.
- Site-based raters administered the QIDS-C16, while the HAMD-17 was remotely administered by telephone by a cohort of independent, calibrated clinicians.
- To evaluate the accuracy of QIDS-C16 scores at screening, the percentage of scores near the inclusion threshold (defined as scores 11-13) was calculated per site.
- Next, to compare site-administered QIDS-C16 versus independently rated HAMD-17, the QIDS-C16 scores were translated into HAMD-17 scores using published comparison guidelines. HAMD-translated QIDS-C16 scores were then subtracted from the original scores, with positive scores indicating higher rating by site raters than independent clinicians.
- Evidence of higher proportion of QIDS-C16 scores around inclusion point at a site, coupled with a positive mean difference score, served as a measure of inclusion bias.

REFERENCES
1. Wium-Andersen, I. K., Vinberg, M., Kessing, L. V., & McIntyre, R. S. (2017). Personalized medicine in psychiatry. Nordic journal of psychiatry, 71(1), 12-19.
2. Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., & Schatzberg, A. F. (2016). Major depressive disorder. Nature Reviews Disease Primers, 2, 16065.
3. Rush AJ, Trivedi MH, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry. 2003;54:573–583.
4. Hamilton M: A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry 23:56-62, 1960.

RESULTS

- Frequency distributions for QIDS-C16 and HAMD-17 total scores at screening and baseline, along with the HAMD-translated scores, are shown in **Figure 1**. The overall means (SD) at baseline were 16.01(2.9) for QIDS-C16 and 20.57 (4.9) for HAMD-17.
- The correlations between original and HAMD-17 translated scores were low at baseline ($r = .45$, $p < .0001$) compared to subsequent visit at week 4 ($r = .70$, $p < .0001$), indicating less agreement at baseline, where inclusion thresholds impacted subject selection.
- At site level, **Figure 2** shows the QIDS-C16 difference scores between screening and baseline, where a positive score indicates higher score at screening than baseline. Most scores were around the axis indicating small differences, with some sites showing very large differences.
- Further, the percentage of scores near inclusion threshold by site (bar graph), along with the site vs. translated mean difference scores (square symbol) was plotted in Figure 3.
- The percentages of assessments at or near inclusion threshold ranged from 0-61 percent, with more than half of the sites having at least 20 percent of their assessments in the 11-13 range for the QIDS-C16.
- The mean difference scores also identified sites with positive scores, suggesting inflation of scores at baseline at some sites. There was, however, little overlap between higher inclusion point percentage and a positive mean difference score, indicating that inclusion bias (i.e., differences between independent ratings and site ratings) was not evident.

Figure 1: Frequency Distributions for QIDS-C16 and HAMD-17 at Screening and Baseline

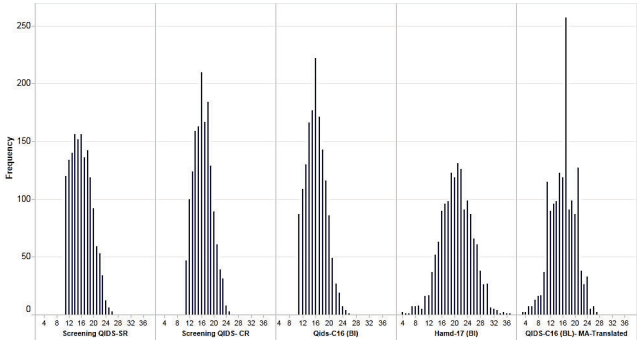


Figure 2: QIDS-C16 Difference Scores by Site (Screening – Baseline)

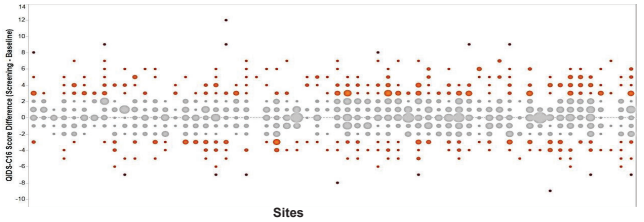
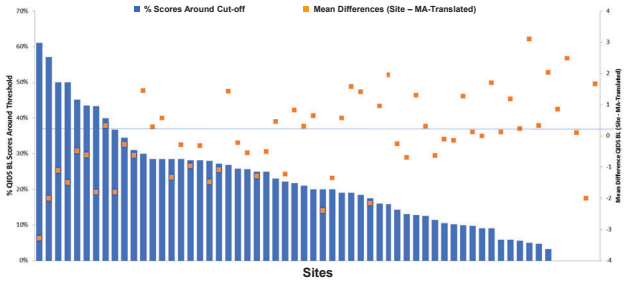


Figure 3: QIDS-C16 Baseline Around Inclusion Threshold with Mean Difference Scores



DISCUSSION

- In this study, sites with high percentages of subjects at or near the inclusion threshold did not show significant evidence of inclusion bias.
- Differences between site-based and independent, remote raters suggest that some patients may have been differentially rated by the two methodologies, but not in a systematic way when evaluated at the site level.
- Efforts to identify and remedy data quality issues and bias remain critical, both to help minimize noise and increase signal detection and to help ensure objectivity of results for the next generation of personalized medicine studies.

MedAvante-ProPhase
A WIRB-Copernicus Group Company