Unexpected Findings: Low Correlation Between Clinician-Administered MADRS Ratings and Patient-Reported HAMD Scores in a Clinical Trial for Major Depressive Disorder

Christian Yavorsky, Cynthia McNamara, Brian K. Saxby, Kristy Wolanski, Francisco Burger, Guillermo Di Clemente

Cronos Clinical Consulting Services, Inc.

Introduction

The Montgomery-Åsberg Depression Rating Scale (MADRS) and the Hamilton Depression rating scale (HAMD) are the most widely used assessments in clinical trials for depression. Many studies (e.g., Jiang & Ahmed, 2009) indicate moderate (r=0.62) to strong (r=0.92) correlations when these scales are administered by trained clinicians. Strong correlation (r=0.96) between patient and clinician ratings has also been reported for the HAMD (Kobak et al, 2000). Whether these instruments remain correlated when one is clinician-administered and the other is patient report has been shown in academic studies but has not been investigated sufficiently in clinical trials with much larger sample sizes.

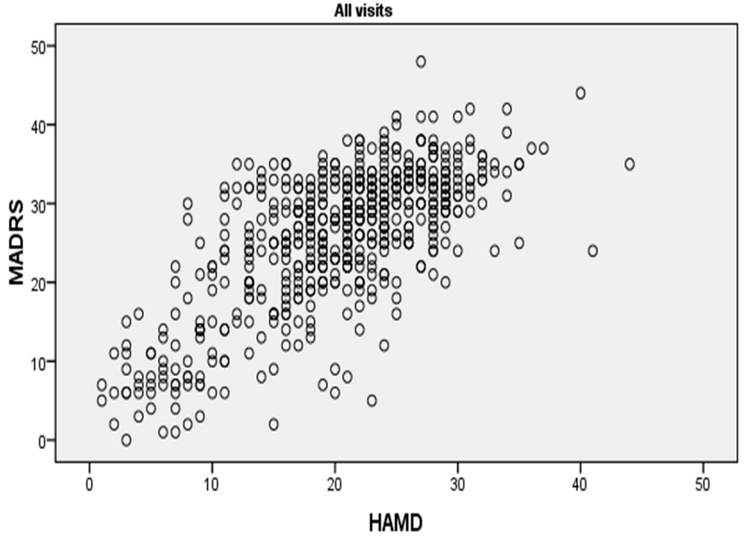
Methods

A randomized, placebo-controlled trial of a compound used to treat patients with Major Depressive Disorder used psychometric data-monitoring to minimize risk associated with human error in study measurement. Computational algorithms identified risk based on predictive analytics (accumulated trial data) alongside scale dynamics (e.g., do items agree). We reviewed data generated by a patient self-report HAMD using Interactive Voice Response (IVR) and the clinician-administered MADRS. The correlation between these scales was used as a potential proxy for risk of assessment error.

Patients with an unsatisfactory response to an adequate trial of an antidepressant medication were entered into a 6-week treatment augmentation study. At each visit, patients completed the 17-item version of the HAMD via Interactive Voice Response (IVR) and trained clinicians administered the MADRS. Eligibility was determined based on the HAMD score at visits 1 and 2, and sites were blinded to the threshold required for study entry. Data from 575 individual patient visits were analyzed using SAS 9.3 and Spearman's correlations obtained between MADRS and HAMD scores across visits. Individual item scores were compared to determine if closely-related symptom domains were being scored consistently across scales.

Results

MADRS and HAMD total scores by visit were moderately correlated at baseline (r=0.454, p<.0001) with weak correlations at visit 2 (randomization visit; r=0.383, p<.0001) and moderate to strong correlations at visits 4, 5 and 6 (scales were not performed at visit 3). The mean MADRS score at baseline was 30.94 (SD 4.52) and the mean HAMD at baseline was 24.48 (SD 5.12). Individual item correlations were weak and not significant between MADRS item "reported sadness" and HAMD item "depressed mood" (r=0.055, p=0.590) at the baseline visit. The two items were more closely correlated by visit 4 (r=0.518, p<.0001). MADRS item "inner tension" and HAMD item "psychic anxiety" had similarly weak correlations at baseline and Visit 2, though the items across scales assessing both sleep and appetite had moderate to strong correlations across all visits.



www.cronosccs.com

Conclusion

Some researchers (e.g., Kobak, 2000) indicate that IVR use of the scale is essentially equivalent to clinician administration. However, we found the two most commonly-used depression scales, the MADRS and the HAMD, did not agree strongly at key study visits when administered by a clinician and completed as patient self-report, respectively. Individual item correlations thought to measure similar constructs across scales had weak correlations, and although they improved as the study continued this may have been an effect of reduced symptoms in response to treatment.

There is also some evidence that patients tend to overestimate depression severity and have limited agreement with the clinician-reported version of the same scale (Kunugi et al, 2013). In our case it appeared that the raters estimated depression severity as consistently higher; baseline and Visit 2 time points seemed especially discrepant between patient and rater assessments.

These results may have implications future trial design, particularly for the assessment of 'fast-acting' antidepressants, as the industry moves towards more at-home or remote assessment methodologies.

References

Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960; 23: 56–62.

Jiang Q, Ahmed S. An analysis of correlations among four outcome scales employed in clinical trials of patients with major depressive disorder. Ann Gen Psychiatry. 2009; 8(4): 1-6.

Kobak KA, Mundt JC, Greist JH, Katzelnick DJ, Jefferson JW. Computer assessment of depression: Automating the Hamilton Depression Rating Scale. Drug Inf J. 2000; 34: 145-156.

Kunugi H, Koga N, Hashikura M, Noda T, Shimizu Y, Kobayashi T, Yamanaka J, Kanemoto N, Higuchi T. Validation of computer-administered clinical rating scale: Hamilton Depression Rating Scale assessment with Interactive Voice Response technology – Japanese version. Psychiat Clin Neuros. 2013; 67: 253-258.

Montgomery S, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 2004; 161: 2163-77.

Disclosures: CY, CM, BKS, KW, FB are employees, and GDC is the Founder & President of Cronos Clinical Consulting Services, Inc., provider of risk-based data monitoring of clinical outcomes and related services to pharmaceutical trials.

Corresponding Author: *Cynthia McNamara, PhD: cynthia.mcnamara@cronosccs.com*

Presented at ASCP 2019 Annual Meeting, May 28-31 2019, Scottsdale, AZ