Impact of Rater Change on Data Variability in MADRS Total Score in Depression Trials Using Site Versus Independent Raters

Echevarria, B[1]., Welch, M[1]., Negash, S[1]., Poppe, C[1]., and Opler, M[1]

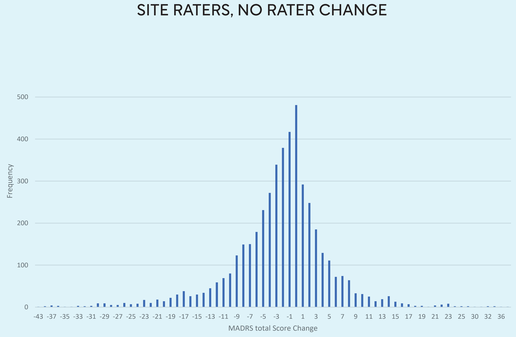
[1]WCG, Princeton, NJ, USA

Investigate how visit-to-visit rater change affects data variability in depression trials using site versus independent (central) raters.

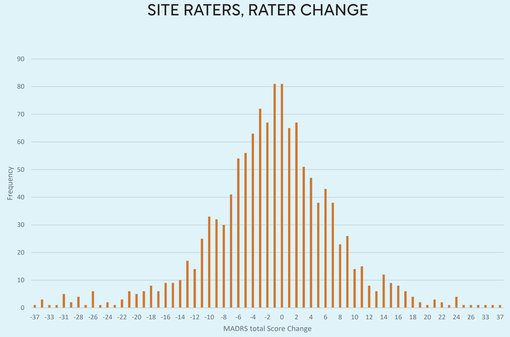
Introduction (Aims)

Depression trials are prone to failure, in part due to noise in endpoint data which may stem from rater factors, including low interrater reliability, poor assessment quality, and rater bias ([1, 2]). Central (Independent) ratings have been proposed as a potential solution to mitigate these issues ([3]). Independent Raters (IR) refer to a group of rigorously trained and tightly calibrated clinicians who are independent to the study sites and blinded to the protocol and study visit. The potential advantages of using Independent Raters in clinical trials include ensuring standardized administration of endpoints, decreased baseline score inflation ([3, 4]), and reduced expectancy bias. In addition to using IR, use of the same rater at each visit has been typically recommended in CTs to attempt to reduce error variance in endpoint data. This study evaluates the impact of rater change in visit-to-visit score variability on MADRS total score in clinical trials of depression using independent raters and site-based raters.

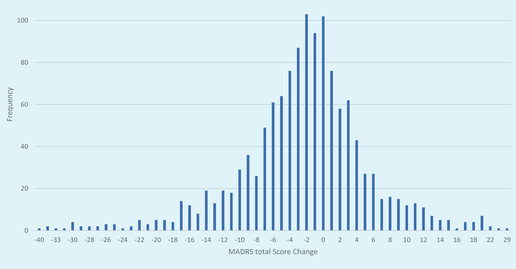
SITE RATERS, NO RATER CHANGE



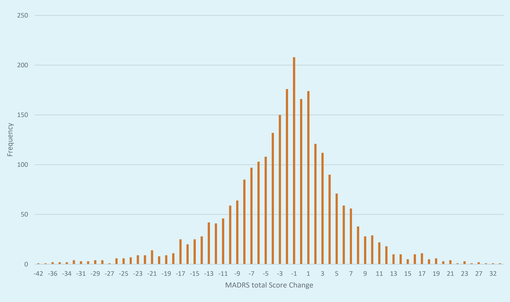
SITE RATERS, RATER CHANGE



INDEPENDENT RATERS, NO RATER CHANGE



INDEPENDENT RATERS, RATER CHANGE



Methods

Data from 20 multi-site double-blind, placebo controlled clinical trials of moderate depression were analyzed. MADRS assessments conducted by site raters (SR) and independent raters (IR) around the world were arranged according to visit sequence and divided into two groups: Rater Change (different raters administered subject's consecutive visits) and No Rater Change (same rater administered consecutive visits). Separate analyses for SR change/no change and IR change/no change were conducted. The change/no change groups were matched on mean interval times between visits using random sampling. Visit-to-visit absolute score changes for MADRS total score were calculated for both cohorts, and frequency distributions were evaluated.

MEAN MAGNITUDE OF DIFFERENCE FOR SITE RATERS

MEAN MAGNITUDE OF DIFFERENCE FOR INDEPENDENT RATERS

Results

A two-way ANOVA with F tests was conducted to compare score changes between the rater change groups. The model controlled for the different visits corresponding with each change score. The MADRS total score, the SR Change group showed a higher mean visit-to-visit score change (N = 1252, mean = 6.53, SD = 6.21) compared to the SR No Rater Change (N = 4673, mean = 5.35, SD = 5.78) group. This group difference reached statistical significance ($F_{1,7}$ = 8.04, p-value < 0.01). On the other hand, the IR change group (N = 2570, mean = 6.08, SD = 5.94) and IR No Rater Change group (N = 1287, mean 5.82, SD = 5.79) did not differ significantly in mean visit-to visit score change ($F_{1,7}$ = 0.11, p-value = 0.74). Rater Change and No Rater Change groups did not differ significantly on mean interval times (SR: t = -0.83, df = 1952, p-value = 0.40; IR: t = 0.12, df = 2495, p-value = 0.91).

Conclusions

Findings from this study indicate significantly higher visit-to-visit score changes when there is a rater change between visits than not for MADRS total score in the SR group. In the IR group, visit-to-visit data variability as result of rater change was not significant. Highest mean absolute score changes occurred from visit[3] to visit[4], immediately post-randomization, in both groups. These results may suggest the importance of maintaining rater consistency in site-based assessments to reduce unwanted variability in MADRS data, while consistency may be less important when using IR. Future studies should aim to improve on this pilot effort and attempt to evaluate the impact of rater consistency on treatment effect using unblinded study data. Additional analyses to further investigate the baseline-to-next visit score change in both groups should be conducted.

References

Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? J Clin Psychopharmacol. 2007 Feb;27(1):1-5. doi: 10.1097/JCP.0b013e31802eb4b7.

Kobak, Kenneth A. PhD; Brown, Brianne PsyD; Sharp, Ian PhD; Levy-Mack, Hollie MSW; Wells, Kurrie PhD; Ockun, Felice MS, MSW; Williams, Janet B.W. DSW. Sources of Unreliability in Depression Ratings. Journal of Clinical Psychopharmacology 29(1):p 82-85, February 2009. | DOI: 10.1097/JCP.0b013e318192e4d7

Kobak KA, Leuchter A, DeBrota D, Engelhardt N, Williams JB, Cook IA, Leon AC, Alpert J. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol. 2010 Apr;30(2):193-7. doi: 10.1097/JCP.0b013e3181d20912.

Williams JB, Kobak KA, Giller E, Reasner DS, Curry L, Detke MJ. Comparison of Site-Based Versus Central Ratings in a Study of Generalized Anxiety Disorder. J Clin Psychopharmacol. 2015 Dec;35(6):654-60. doi: 10.1097/JCP.0000000000000422.