Presented at the 22nd Annual ISCTM Scientific Meeting, Feb 18 - 20, 2026, Washington, D.C.

Comparing Baseline MADRS Distributions by Inclusionary Scale Use in Major Depression Studies

Alan Kott[1]; Christine zu Eulenburg[2]; Xingmei Wang[1]; Andrei Iacob[1]; Hans Eriksson[2]; Gary Sachs[1]

1) Signant Health, 2) HMNC

METHODOLOGICAL ISSUE BEING ADDRESSED

Do baseline MADRS scores differ between major depressive disorder (MDD) trials that use the MADRS as an inclusion criterion and those that do not?

INTRODUCTION

- Accurate assessment of baseline depressive severity is crucial in clinical trials of major depressive disorder (MDD).
- The Montgomery–Åsberg Depression Rating Scale (MADRS) and the Hamilton Depression Rating Scale (HAM-D) are among the most commonly used instruments for this purpose.
- Methodological analyses have highlighted that when a scale is used both as an inclusion criterion and as the primary outcome, knowledge of the entry threshold can promote baseline score inflation, particularly among patients whose symptom severity lies near the cut-off, thereby increasing placebo response and reducing signal detection.
- While this concern is widely acknowledged, empirical, MADRS-specific evidence quantifying the magnitude of such inflation has been limited.
- The present post hoc analysis addresses this gap by comparing baseline MADRS score distributions in MDD trials according to whether MADRS was used for participant selection.

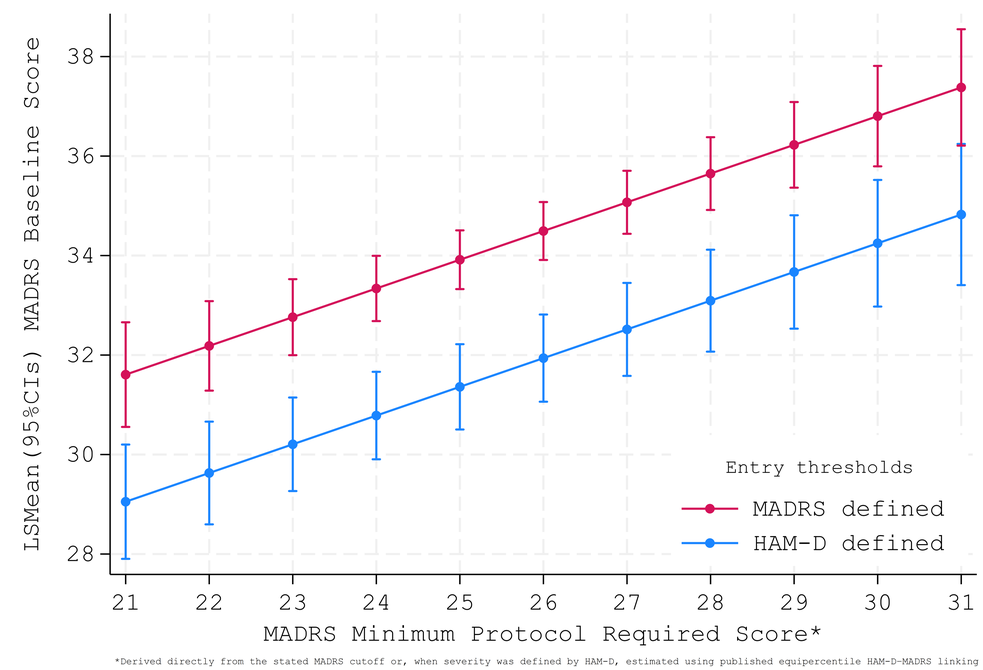
METHODS

- Baseline MADRS data were obtained from 20 double blind, placebo controlled, multicentric clinical trials in MDD.
- A mixed-effects model using restricted maximum likelihood estimation was applied to compare baseline MADRS scores between protocols requiring a predefined MADRS threshold and those requiring a predefined HAM-D threshold.
- Severity thresholds were standardized by adjusting for each protocol's minimum required score, derived directly from the stated MADRS cutoff or, when severity was defined by HAM-D, estimated using published equipercentile HAM-D–MADRS linking (Leucht, 2018).
- Protocol was included as a random effect to account for study-level clustering.

RESULTS

- **The dataset included 7,034 baseline MADRS assessments from 20 clinical trial protocols:**
  - 14 using MADRS and
  - 6 using HAM-D as inclusionary scales.
- **Mean baseline MADRS scores were 31.0 in trials not using MADRS for eligibility and 34.3 in those that did.**
- **The mixed-effects model indicated a statistically significant mean difference of 2.6 points ($p < 0.001$), with higher scores observed in studies where MADRS determined inclusion.**

FIGURE 1: BASELINE MADRS MEAN SCORES BY INCLUSION THRESHOLD DEFINED ON MADRS VS HAM-D SCALES



*Derived directly from the stated MADRS cutoff or, when severity was defined by HAM-D, estimated using published equipercentile HAM-D–MADRS linking

DISCUSSION

- This analysis found that baseline MADRS scores were significantly higher in trials using the scale as an inclusion criterion compared with those relying on HAM-D thresholds, with an adjusted mean difference of approximately 2.6 points.
- Because the model controlled for protocol-specific cut-offs, the observed difference cannot be explained by inherently stricter inclusion criteria.
- Instead, the pattern aligns with longstanding concerns that when the same scale is used for eligibility and outcome assessment, raters who are aware of the inclusion threshold may consciously or unconsciously inflate baseline scores, especially for patients near the cut-off, to secure trial entry.
- Such inflation can distort baseline severity, impact symptom change in both treatment arms, and possibly attenuate drug–placebo differences.
- These findings complement prior methodological recommendations to separate screening and outcome measures or to use independent/central raters in order to reduce baseline inflation and improve signal detection in MDD trials.
- A limitation of this analysis lies in the absence of a corresponding HAM-D evaluation, which would clarify whether a similar effect occurs when HAM-D serves as the inclusionary measure.

Disclosure: Authors declare no conflicts of interest.

SIGNANT HEALTH
THE EVIDENCE GENERATION COMPANY