Data Patterns Predict MADRS-CGI Discrepancies in Depression Trials: Implications for Rater Training and Data Monitoring

Presenters: Karazsia, Bryan T; Engler Jenicka; Sirbu, Cristian, Patrick, Katrina

All Authors/Presenters are affiliated with Cronos Clinical Consulting Services, Inc. an IQVIA Business

Introduction

Two of the most commonly utilized Clinician Reported Outcomes (ClinROs) in antidepression trials (ADTs) are the:

- Montgomery-Åsberg Depression Rating Scale (MADRS)[1]
- Clinical Global Impressions-Severity (CGI-S) scale[2]

Prior research indicates these two measures are substantially interrelated; in the context of clinical trials, the following has been demonstrated:

- Strong agreement across identification of 'responders' and 'non-responders' to treatment[3]
- Statistically significant correlations of sufficient magnitude to justify 'linking analysis' (i.e., a method for translating scores from one measure to another)[4]

Due to these strong interrelations, discrepancies between MADRS and CGI-S are important to monitor:

MADRS-CGI-S Discrepancy: Scores for same participant at same visit that reflect a different degree of severity across the MADRS and CGI-S

Why might discrepancies occur?

- Although the measures are highly correlated, though they are not 100% interchangeable
- CGI may capture aspects of participant presentation not reflected in the 10 items of the MADRS, such as unique cultural presentations[5]
- Possible errors, including:

  Rater Scoring Error: Rater selects a score that does not align with participant presentation

  Data Entry Error: Scores are captured in EDC that do not reflect score assigned by Rater

Purpose of Present Study:

1) Examine frequency of MADRS-CGI-S Discrepancies in global Major Depressive Disorder (MDD) clinical trials
2) Identify predictors of MADRS-CGI-S Discrepancies

Methods

Data Source: Merged data across 5 global Major Depressive Disorder clinical trials that included the MADRS and CGI-S at two timepoints (N = 1722; Baseline & End of Treatment [EoT] visits)

Using Baseline data (100% of participants) and available EoT data (81.2% of participants), there were a total of 3120 MADRS-CGI scoring comparisons available for analysis

Variables used for analysis:

- **MADRS-CGI Discrepancy Operationalization:** 0/1 dichotomy based on published scoring translations[4]
- MADRS individual item scores
- Bivariate interactions of MADRS items
- Timepoint: Baseline & EoT

We split the data into two random subsamples (n = 1560) to allow a dataset for model testing and a second dataset for validation

Analytic Plan

Logistic regression models predicting MADRS-CGI Discrepancy

Sample 1: Due to the number of possible predictors (10 items, 2 timepoints, 66 possible interactions) we used Lasso[6] within a logistical regression framework to reduce overfitting and optimize variable selection

Sample 2: Logistic regression using variables identified in analysis in Sample 1 (See Tables 1a an 1b).

Results

Frequencies of Discrepancies: **Sample 1**: 33 identified discrepancies (**2.12%**); **Sample 2**: 20 identified discrepancies (**1.28%**)

Tables 1a & 1b: Logistic regression analyses predicting scoring discrepancies

Table 1a: Validation of LASSO-selected logistic regression model in Sample 2

Predictors	Coef.	SE	*p*	$CI_{95\%}$
MADRS: Reported Sadness	.02	.16	.90	-.30 - .34
MADRS: Apparent Sadness	.03	.14	.83	-.25 - .31
MADRS: Inner Tension	.00	.12	.99	-.23 - .24
MADRS: Reduced Appetite	-.04	.16	.79	-.35 - .27
MADRS: Inability to Feel	.04	.31	.89	-.23 - .12
MADRS: Suicidal Thoughts	1.37	.73	.06	-.57 - .65
Reduced Sleep x Suicidal Thoughts	.02	.07	.77	-.07 - 2.81
Apparent Sadness x Reduced Appetite	.00	.04	.98	-.11 - .15
Apparent Sadness x Lassitude	.04	.02	.06	-.07 - .07
Inner Tension x Reduced Appetite	.01	.03	.66	-.00 - .09
Reduced Sleep x Reduced Appetite	**.06**	**.02**	**.003**	**.02 - .10**
Concentration Difficulties x Pessimistic Thoughts	**.05**	**.15**	**<.001**	**.02 - .08**
Visit x Reported Sadness	-.18	.18	.31	-.54 - .17
Visit x Item 7	-.07	.11	.53	-.30 - .16

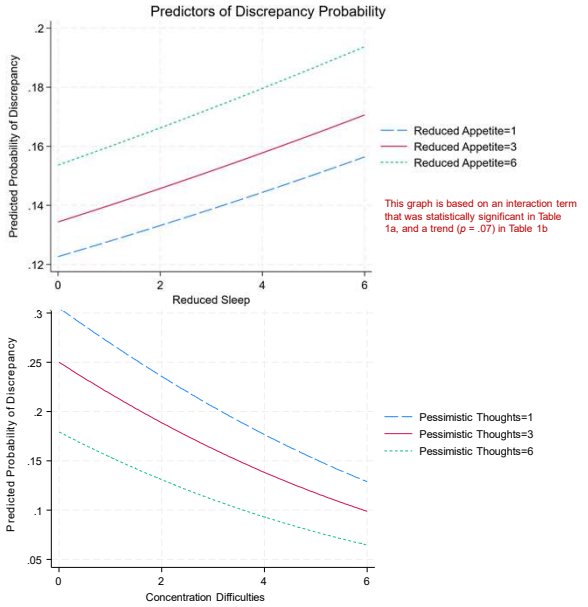
LR $\chi^2(15)$ = 137.40, p<0.001; Pseudo R^2=0.10; AUC=0.72; Significant predictors are **bolded**.

Table 1b: Truncated model including only significant interaction terms and their associated individual predictors in Sample 2

Predictors	Coef.	SE	*p*	$CI_{95\%}$
MADRS: Reduced Sleep	.05	.09	.57	-.12 - .22
MADRS: Reduced Appetite	.06	.11	.63	-.17 - .28
MADRS: Concentration Difficulties	**-.23**	**.09**	**.01**	**-.41 - -.05**
MADRS: Pessimistic Thoughts	-.18	.12	.14	-.41 - .06
Reduced Sleep X Reduced Appetite	.05	.03	.07	-.00 - .11
Concentration Difficulties X Pessimistic Thoughts	**.12**	**.03**	**.00**	**.06 - .19**

LR $\chi^2(6)$ = 122.08, p<0.001; Pseudo R^2=0.09; AUC=0.71; Significant predictors are **bolded**.

Figures: Visualizations of Significant Interaction Terms (based on Model portrayed in Table 1b)



Conclusions

The present study illustrated that

1) MADRS-CGI discrepancies can and do occur during trials.
2) MADRS-related variables increase predicted probability of a discrepancy.

Across the models, the following variables emerged as significant predictors:

- Interaction: Concentration Difficulties x Pessimistic Thoughts: presence of low scores on both items is associated with a higher predicted probability of a discrepancy the predicted probability of a discrepancy. This was the only construct that emerged as a significant predictor in the 2 models tested in Sample 2.
- Interaction: Reduced Sleep x Reduced Appetite: profile with high scores on both items is associated with a higher predicted probability of a discrepancy. This construct was not significant in the truncated model.
- Concentration Difficulties: The only single item to emerge as a significant predictor, though only in the truncated model, so we advise caution when interpreting.

Implications for Rater Training: Raters should always ensure that scores they assign reflect participant presentations; results from this study suggest that some scoring patterns warrant reflection on the data before final submission. When the scoring patterns identified here occur in trials,

- Raters may want to review data across scales (MADRS and CGI) and verify that values recorded represent the functioning of the participant.
- Study teams may wish to ensure that Rater scores represent their intended ratings and participant presentations.

References

1. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382-389.
2. Guy W, ed. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: US Dept of Health, Education, and Welfare; 1976.
3. Morrens J, Mathews M, Popova V, Borentain S, Rive B, Gonzalez MMB, Jamieson C, Zhang Q. Use of the Clinical Global Impressions-Severity (CGI-S) to assess response to antidepressant treatment in patients with treatment-resistant depression. *Neuropsychiatr Dis Treat*. 2022;18:1127-1132.
4. Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Lepping P, Szegedi A. What does the MADRS mean? Equipercentile linking with the CGI using a company database of mirtazapine studies. *J Affect Disord*. 2017;210:287-293.
5. Engler J. Considerations for risk-based data monitoring and eligibility review in depression global trials. Presented at: ISCTM Autumn Conference; 2025;Amsterdam,Netherlands.
6. StataCorp. *Stata 19 Lasso Reference Manual*. College Station, TX: Stata Press; 2025.

Disclosures: All authors report no conflicts of interest for this work. All are current employees of Cronos Clinical Consulting Services, Inc. an IQVIA Business. Cronos does provide Rater training, Rater oversight and Rater intervention services.

Cronos

Cronos Clinical Consulting Services, Inc. an IQVIA Business