



# Reliability and Validity of Clinician and Computer-Administered MADRS Assessed in Randomized Controlled Trials

Robert O. Knauz, PhD<sup>1,2</sup>, Daniel DeBonis<sup>1</sup>, Gary S. Sachs, MD<sup>1,2</sup>

1. Concordant Rater Systems, Boston, MA 2. Massachusetts General Hospital

## Introduction

Computer-administered assessments such as the Montgomery Asberg Depression Rating Scale (MADRS) have been used in double-blind, placebo controlled studies as primary assessments (an alternative to site based human raters) and as part of a quality control process. (in tandem same scale administration by site based human raters.)

The acceptability of these techniques rests largely on studies conducted to specifically determine their reliability and validity. Since studies done for the purpose of establishing reliability and validity may over estimate performance in circumstances of routine use, there is a need to examine measures of performance in actual controlled randomized trials.

Since 2003, Concordant Rater Systems (CRS) has administered computerized assessments in tandem with site based clinician administered assessments. At this time, 2685 pairs of MADRS assessments from US and International trials in 7 languages have been analyzed.

CRS data sets offer an opportunity to evaluate the performance of site based raters in comparison to computer-administered assessments at specific time points of relevance to clinical trialists engaged in drug development: Baseline, Post-randomization, and Study Endpoint.

## Methods

Available CRS datasets from randomized, controlled studies conducted between January 2003 and November 2008 were reviewed to identify those with MADRS scores from site based raters (MADRS<sub>SBR</sub>) and computer-administered assessments (MADRS<sub>Comp</sub>). The datasets had no subject identifiers other than the subject's study identification number.

The sample included 2685 pairs of MADRS<sub>SBR</sub> and MADRS<sub>Comp</sub> ratings. Internal scale consistencies were calculated using Cronbach's alpha, mean item and total score correlation.

Correlations between the pairs were calculated to examine measurement reliability overall and individually for each MADRS item. Comparisons of the MADRS<sub>SBR</sub> and MADRS<sub>Comp</sub> were made at Baseline, Post-randomization, and Study Endpoint.

All ratings were made in the subject's native language. Raters with high discordance received remediation by telephone during the course of the study.

## Results

Figure 1: Interclass Correlation Coefficients N=2685 MADRS<sub>SBR</sub> - MADRS<sub>Comp</sub> pairs

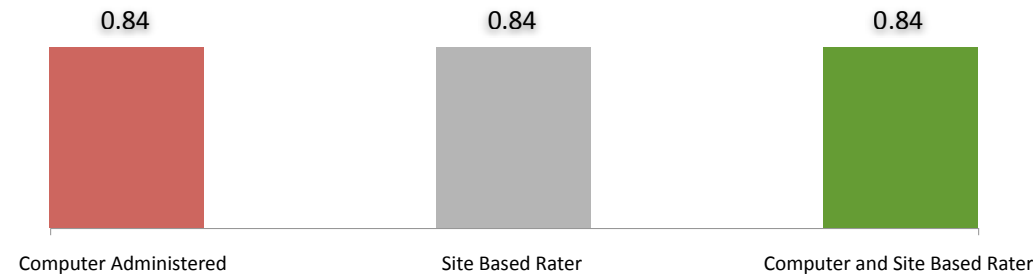


Figure 2: ICC N=2685 MADRS<sub>SBR</sub> - MADRS<sub>Comp</sub> pairs

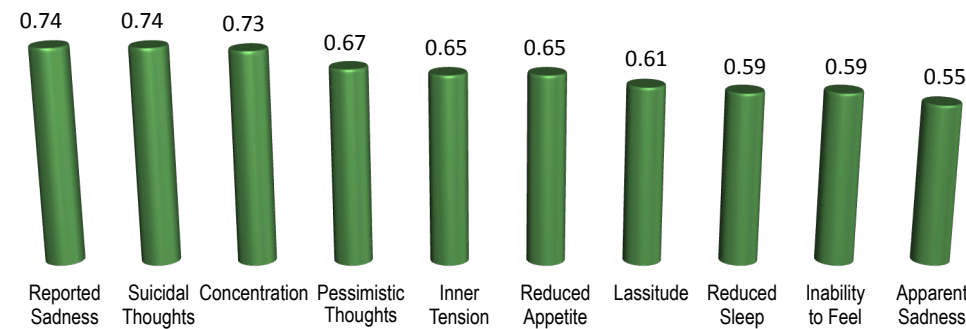
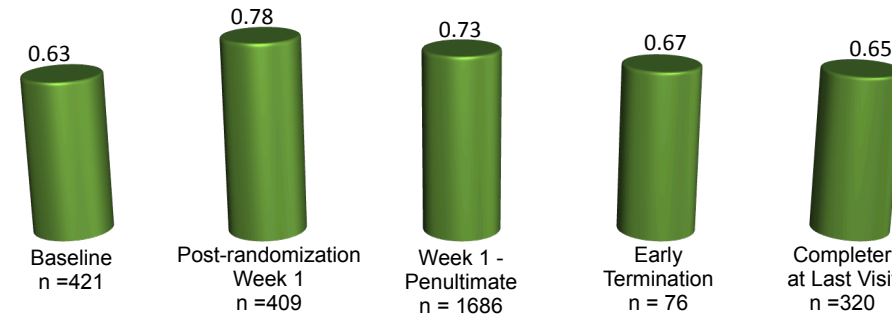


Figure 3: MADRS ICC at Specified Study visit



Analysis funded by Concordant Rater Systems, Boston, MA  
617-367-0044 | www.concordantrater.com

Figure 4: Total Scores for paired MADRS<sub>SBR</sub> and MADRS<sub>Comp</sub> conducted in subject's native language at specified study visit

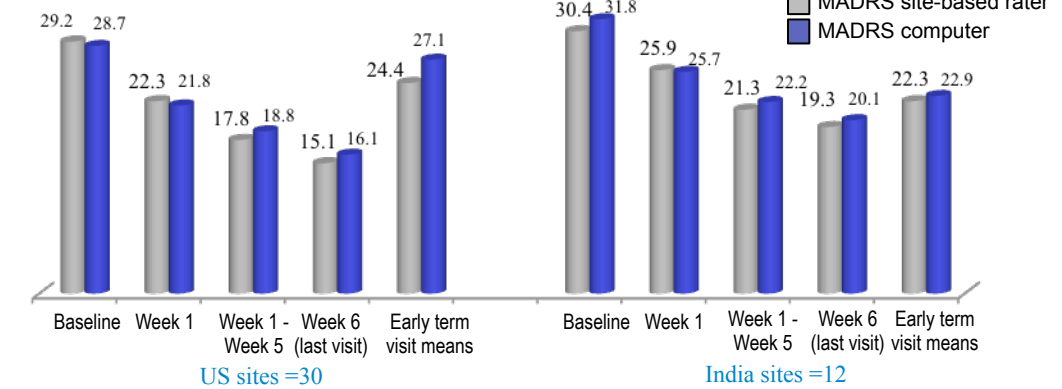
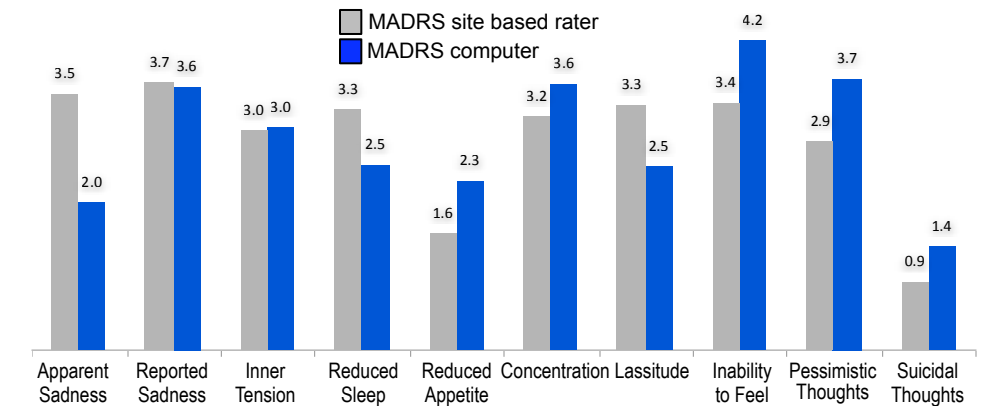


Figure 5: Mean Item Scores at Baseline for 421 paired MADRS<sub>SBR</sub> and MADRS<sub>Comp</sub>



## Conclusions

1. Close agreement between MADRS<sub>Comp</sub> and MADRS<sub>SBR</sub> across conditions and cultures during the conduct of actual clinical trials supports the validity and reliability of computer-administered MADRS.
2. The MADRS<sub>Comp</sub> is a valid measure that appears to be as sensitive to change as the MADRS<sub>SBR</sub>.
3. MADRS<sub>Comp</sub> may be a useful benchmark against which the performance of site based raters in global multisite studies can be measured. Identification and remediation of poor performing site based raters may reduce hazards associated with failed studies.