

Independent Telephone-based Assessment of Depressive Symptoms in China

Gang WANG¹, Tianmei SI², Xiufeng XU³, Jaskaran SINGH⁴, Jessie HE⁵, Cong LIU⁶, Lawrence H. Yang^{7,8}, Honglan LI⁴, Jianmin ZHUO⁵, Mark Opler^{6,9}

¹Beijing Anding Hospital of Capital Medical U.; ²Peking U. Sixth Hospital; ³First Affiliated Hospital of Kunming Medical U.; ⁴Janssen R&D; ⁵Janssen China R&D Center; ⁶MedAvante-ProPhase, Inc.; ⁷NYU College of Global Public Health; ⁸Mailman School of Public Health, Columbia U.; ⁹New York U. Medical Center

METHODOLOGICAL QUESTION:

Is the MADRS reliable when administered via telephone to Mandarin-speaking patients with Major Depression Disorder in China?

INTRODUCTION

- A pilot study was conducted at three hospitals in China to study the use of telephone administration of the Structured Interview Guide for the Montgomery-Asberg Depression Rating Scale (SIGMA).
- The Montgomery-Asberg Depression Rating Scale (MADRS) is a 10-item clinician-administered scale designed to be used in subjects with MDD to measure the overall severity of depressive symptoms¹.
- The MADRS scale was selected as the primary efficacy measure for this study because it is validated, reliable², and acceptable to regulatory health authorities as a primary scale to determine efficacy in major depression.
- The primary objective for this study is to collect preliminary reliability and validity data for the MADRS as administered by telephone in a sample of Mandarin-speaking patients diagnosed with MDD.



METHODS

- Five subjects were recruited at each site for a total of 15 subjects.
- Each subject was assessed at three visits (Screening, Week1, Week 2) by a site-based rater (SBR) who completed an in-person SIGMA interview and an independent rater (IR) who completed the SIGMA interview by telephone. The SBR also completed the CGI at each visit.
- Rater reliability was calculated using standard methods (e.g., intra-class coefficient)
- We modeled the means across time and the mean difference between raters at each time point using a linear mixed model with main effects for Visit and Rater Type as well as interaction effect and random effect.
- Total scores were analyzed using linear mixed-effects regression, LMER, with the lme4 package in R³ fit by maximum likelihood t-tests use Satterthwaite approximations to degrees of freedom.
- To evaluate item-level data, we calculated the mean difference of IR and SBR ratings by item.

RESULTS

- The overall rater reliability between site-based and independent telephone ratings was high.

	Screening	Week 1	Week 2
ICC	0.93	0.82	0.82
- Results also demonstrate a high level of reliability across visits and items in both IR and SBR Chinese clinicians.
- The IR raters tended to rate subjects higher than the site-based raters for the same visit. (Figure 3)
- Visit had little impact on the difference in ratings as the mean MADRS scores for SBRs were consistently lower than IRs across visit. (Figure 1)
- Subjects with low symptom severity (average MADRS Total less 16, or CGI category is 2-3) were more likely to have a higher rating by the IR than the SBR. (Figure 2)
- The output of the linear mixed model shown the main effect of rater type was not statistically significant ($p=0.40$) at Screening in our analysis, neither was the interaction effect ($p=0.25$ for week 1 SBR and $p=0.277$ for week 2 SBR)
- Item-level differences within the data were generally low ($\pm <0.5$ points/item), further reflecting the strong levels of reliability overall. (Table1)

FIGURE 1: Mean of Total Score by Visit

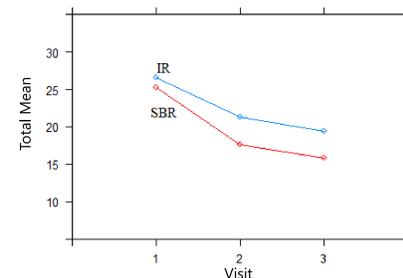


FIGURE 2: Mean of Total Score by CGI Category

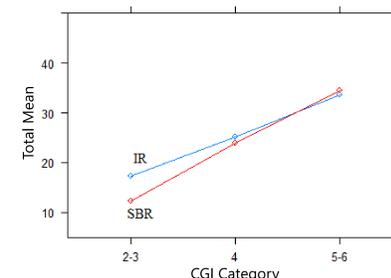


FIGURE 3: IR MADRS Total Score vs SBR MADRS Total Score

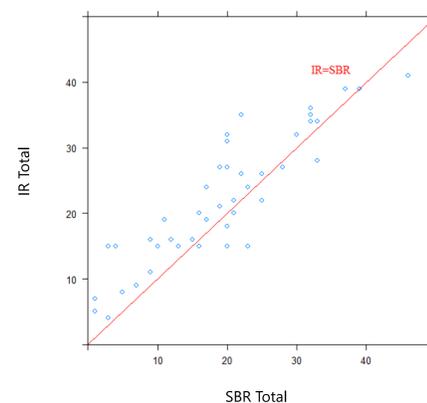


TABLE 1: Item-level differences

ITEM	Apparent Sadness	Reported Sadness	Inner Tension	Reduced Sleep	Reduced Appetite
Difference	0.53	0.31	0.58	0.24	0.33
ITEM	Concentration Difficulties	Lassitude	Inability to Feel	Pessimistic Thoughts	Suicidal Thoughts
Difference	0.31	0.18	0.13	0.16	0.09

CONCLUSION

- In this study, many subjects in treatment demonstrated a decrease in overall symptom over the three study visits. While IR and SBR scores were highly correlated, independent raters tended to rate subjects modestly higher at the same visit than the site-based raters, especially when subjects demonstrated lower overall symptom severity.
- Despite the difference in mean scores, there is not enough evidence to support an alternate hypothesis of a difference in ratings between telephone raters and in-person raters.
- Several of the items cited in other studies as being potentially problematic via telephone (e.g., inability to feel, pessimistic thoughts, and suicidal thoughts – Items 8-10⁴) show the lowest mean differences overall. This may be due to a willingness on the part of patients in this study to more openly discuss and describe these phenomena than has been previously reported.



References:

- ¹ Furukawa, T. A. (2010). Assessment of mood: guides for clinicians. *Journal of psychosomatic research*, 68(6), 581-589.
- ² McDowell, I. (2006). *Measuring health: a guide to rating scales and questionnaires*. Oxford University Press, USA.
- ³ Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, 1(7), 1-23.
- ⁴ Mundt, J. C., Katzelnick, D. J., Kennedy, S. H., Eisfeld, B. S., Bouffard, B. B., & Greist, J. H. (2006). Validation of an IVRS version of the MADRS. *Journal of Psychiatric Research*, 40(3), 243-246.

MedAvante-ProPhase

A WIRB-Copernicus Group Company