eCOA Prompted MADRS Interview: Balancing Thorough Assessment and Efficiency

Roy, M[1], Brown, J[2], Hong, B[2], Benecke, RM[2], Tackett, ZC[3], Zhou, W[2], Barbone, JM[2], Feaster, T[2], Sachs, G[4,5]

[1] Signant Health, Prague, CR, [2] Signant Health, Blue Bell, PA, USA, [3] Signant Health, Arlington, VA, USA, [4] Signant Health, Boston, MA, USA, [5] Massachusetts General Hospital, Boston, MA, USA

BACKGROUND

The Montgomery–Asberg Depression Rating Scale (MADRS)[1] is one of the established, gold standard tools for clinical trial evaluation of depressive treatment efficacy, and has often been used to guide appropriate study subject selection. The Structured Interview Guide for MADRS (SIGMA) published in 1988[2], is a paper prompted interview (PPI) that provides clinicians with suggested questions to use in gathering sufficient information for clinician rating of the 10 MADRS items.

The concept of tandem ratings was later established as a means of evaluating the accuracy of site investigator ratings. With this mission in mind, the electronic Computer Prompted Structured Interview (CPI) and the Computer Simulated Rater (CSR) ePRO scale were developed to operate as a pair designed to provide a within-visit mechanism of assessing site rater rating accuracy. The CPI is an electronic, interactive, clinician-administered interview delivered via the Signant Health Rater Station device, and driven by an algorithm that selects the logical next question for the clinician to ask based on the previous response(s) and guides the interview flow via on-screen prompts. Using information gathered from the subject during CPI administration, the investigator ultimately rates the 10 MADRS items.

The aim of this study is to evaluate whether there is a difference in interview duration times between the PPI and CPI in order to better understand if a particular mode of administration is better able to balance thorough clinical assessment with interviewing efficiency.

METHODS

Data were obtained from 6 studies that used different formats of MADRS administration (PPI or CPI delivered on the Rater Station device) to assess severity of current depression symptoms. To account for potential confounds, only data from double-blind, US English visits with a look back period of past 7 days was used. Duration of administration is calculated based on timestamps captured by the Rater Station device, when a rater clicks "Begin Interview" until the rater signs off on the scale. Durations under 1 minute or greater than 120 minutes were eliminated due to a lower confidence in the validity of such timestamps (e.g. back-entry from paper). The data were divided into two groups based on subject depressive severity:

- **More Severe.** (MADRS total score > 30) This data set contained 2,339 visits and of those visits, 1,087 visits (46.5%) used CPI and 1,252 visits (37.6%) used PPI. To reduce bias caused by differences in sample size, up-sampling was performed to increase the group size to the largest group. After up-sampling, there were 2,504 visits with equal percentage distribution in the two types of MADRS administrations.
- **Less Severe.** (MADRS total score ≤ 30) This data set contained 9,566 visits and of those visits, 9,321 used CPI (97.4%) and 245 (1.67%) used PPI. Similar up-sampling was completed, resulting in 18,642 visits with equal distribution in the two types of MADRS administrations.

The Shapiro-Wilk test was used to test the normality of the interview duration by MADRS administration format. Kruskal Wallis non-parametric test was employed to detect if the interview duration between the PPI and CPI differed significantly. A post-hoc Dunn test of multiple comparison was subsequently conducted to see which scale format differences were statistically significant. Statistical significance was defined as $p < 0.05$. All analyses were performed using R v3.6.1.

RESULTS

More Severe (MADRS total score > 30). A Shapiro-Wilk test for normality on interview duration by scale format found the distribution to be significantly different from a normal distribution (Figure 1). A Kruskal Wallis test found the influence of scale format on interview duration to be significant ($p < 0.0001$). A subsequent post-hoc Dunn test PPI vs. CPI comparison to be significant. For More Severe Group, the CPI had a longer mean interview duration of 20.40 minutes and PPI had an interview duration of 17.59 minutes.

Less Severe (MADRS total score < 30). A Shapiro-Wilk test for normality on interview duration by scale format also found the distribution to be significantly different from a normal distribution (Figure 2). A Kruskal Wallis test found the influence of scale format on interview duration to be significant ($p < 0.0001$). A subsequent post-hoc Dunn test found PPI vs. CPI comparison to be significant. For Less Severe Group, the PPI had the longer mean interview duration of 17.61 minutes and CPI had an interview duration of 10.70 minutes (Table 2).

Figure 1. More Severe - Graphical Summary

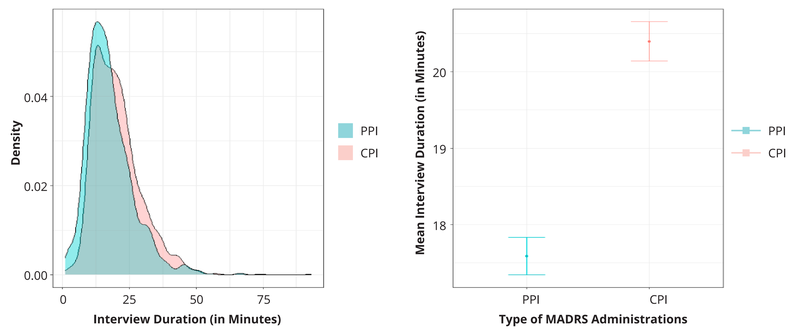

Figure 2. Less Severe - Graphical Summary

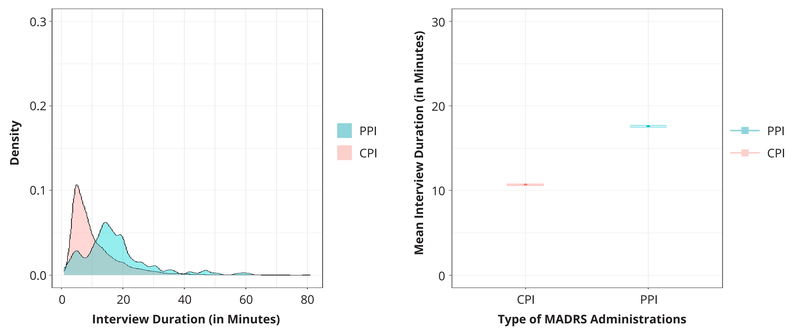

Table 1. More Severe - Summary Table

MADRS Administration	Mean Interview Duration (in Minutes)	Standard Error
CPI	20.40	0.244
PPI	17.59	0.257

Table 2. Less Severe - Summary Table

MADRS Administration	Mean Interview Duration (in Minutes)	Standard Error
CPI	10.70	0.083
PPI	17.61	0.111

DISCUSSION

Results demonstrate a significant difference in interview duration between the MADRS scale administration formats analyzed in the current investigation for both More Severe (MADRS total > 30) and Less Severe (MADRS total < 30). For More Severe subjects, the CPI interview had a longer duration than the PPI. For Less Severe subjects, the CPI had a shorter duration than PPI.

These findings support the value of the CPI for subjects across the severity continuum. Results suggest that its interview format results in more efficient probing with subjects showing relatively less severe symptoms, and drives more thorough rater probing when employed with subjects experiencing particularly severe depressive symptoms. The CPI's thoroughness and efficiency can be attributed to this eClinRO scale's pre-programmed question mapping that encourages investigators to ask all necessary questions to gather the information needed to rate with precision, while also guiding raters to engage in efficient interviewing with subjects who have a relatively lower level of severity.

Future research is suggested to compare the CPI and PPI with regard to rating quality metrics, and if any quality differences are moderated by interview duration. The CPI provides the investigator with optimal assessment questions in a systematic manner that increases the likelihood that the investigator will ask all necessary questions rather than skipping out of key questions at the sacrifice of rating accuracy. Our belief is that this administration mechanism improves rating precision and efficiency in depression trials. Subsequent research should also evaluate if investigator MADRS experience moderates interview format duration differences.

REFERENCES

1. Montgomery SA, Äsberg M. A new depression scale designed to be sensitive to change. Br. J. Psychiatry. 1979;134:382–389.

2. Williams, J.B., & Kobak, K.A. (2008). Development and reliability of a structured interview guide for the Montgomery Asberg Depression Rating Scale (SIGMA). The British journal of psychiatry : the journal of mental science, 192 1, 52-8 .

Financially supported by Signant Health.

Presented at the ISCTM 2020 Autumn Virtual Conference, September 21 - 25, 2020.

SIGNANT HEALTH