

Is a computer simulated rater good enough to administer the Hamilton depression rating scale in clinical trials?

Sachs, G^{1,2}, DeBonis, D¹, Wang, X¹, Epstein, B¹

¹Bracket Wayne, PA, USA ²Massachusetts General Hospital, Boston, MA, USA

METHODOLOGICAL QUESTION

Are Hamilton depression rating Scale scores obtained by a computer simulated rater within the range expected from site-based raters? Concern about the high rate of failed clinical trials fuels the desire for better measurement techniques.

METHODS

- Blinded data was harvested from a double blind placebo controlled industry sponsored study. At each study visit that required the site based rater to administer the HDRS-24, the CSR administered the HDRS-24 as a separate independent rating.
- The CRS conducts an interactive interview directly with the study subject. An interview algorithm selects probe questions based on the subject's last response and a scoring algorithm maps the subject's responses to a unique anchor point.
- Site based raters administering the HAMD were required by the sponsor to have > 2 years of experience with the HAMD 24 and meet training and certification standards based on scoring video tapes viewed at investigator meetings. Raters were granted certification, if the total scores were within 5 points of the expert consensus score and no more than two items were discordant (scores differed by more than 1 point).

INTRODUCTION

- Hamilton depression rating Scale (HDRS) was developed as an instrument which scored multiple symptoms of depression based on information elicited during a face-to-face clinical interview.
- Psychometric data suggests several strategies can improve performance of the HDRS including rater training¹, structured clinical interviews² and explicit definition of anchor points³. These techniques encourage raters to apply a rules based approach to their subjective ratings.
- Unfortunately results from actual randomized clinical trials do not always show the benefit of these techniques (Kahn, ACNP 2014, Sachs CINP 2014) and video certification exercises commonly produce a distribution of total scores with standard deviation >3.5. The evolution of clinical ratings toward a rules based approach encouraged the authors to develop a computer simulated rater (CSR) based on the scripting and rules taught to site-based raters training to administer and score the HDRS-24.

RESULTS

Results were obtained from the Bracket HAMD-24 blinded study dataset which included 737 subjects, 112 raters, and 3180 administrations of the paired rater and computer interviews made over the course of a 16 week double blind placebo controlled clinical trial.

Table 1: HDRS-24 and ICC across study visits

Study Visit	Sample Size	HDRS Mean (\pm s.d.)		p-value	ICC
		SBR	CSR		
Screening	723	29.5 (\pm 5.7)	29.9 (\pm 7.8)	0.2401	0.58
Week 0	510	29.3 (\pm 5.7)	29.3 (\pm 8.4)	0.9266	0.60
Week 2	470	24.1 (\pm 7.7)	23.4 (\pm 9.0)	0.1816	0.73
Week 8	401	18.7 (\pm 9.5)	19.3 (\pm 9.9)	0.3658	0.82
Week 9	371	16.8 (\pm 9.7)	18.2 (\pm 10.2)	0.0547	0.85
Week 12	360	15.5 (\pm 9.4)	16.9 (\pm 10.2)	0.0288	0.84
Week 16	345	13.2 (\pm 9.2)	15.1 (\pm 10.2)	0.0598	0.82

Figure 1: Tandem ratings HDRS-24 total score across study visits

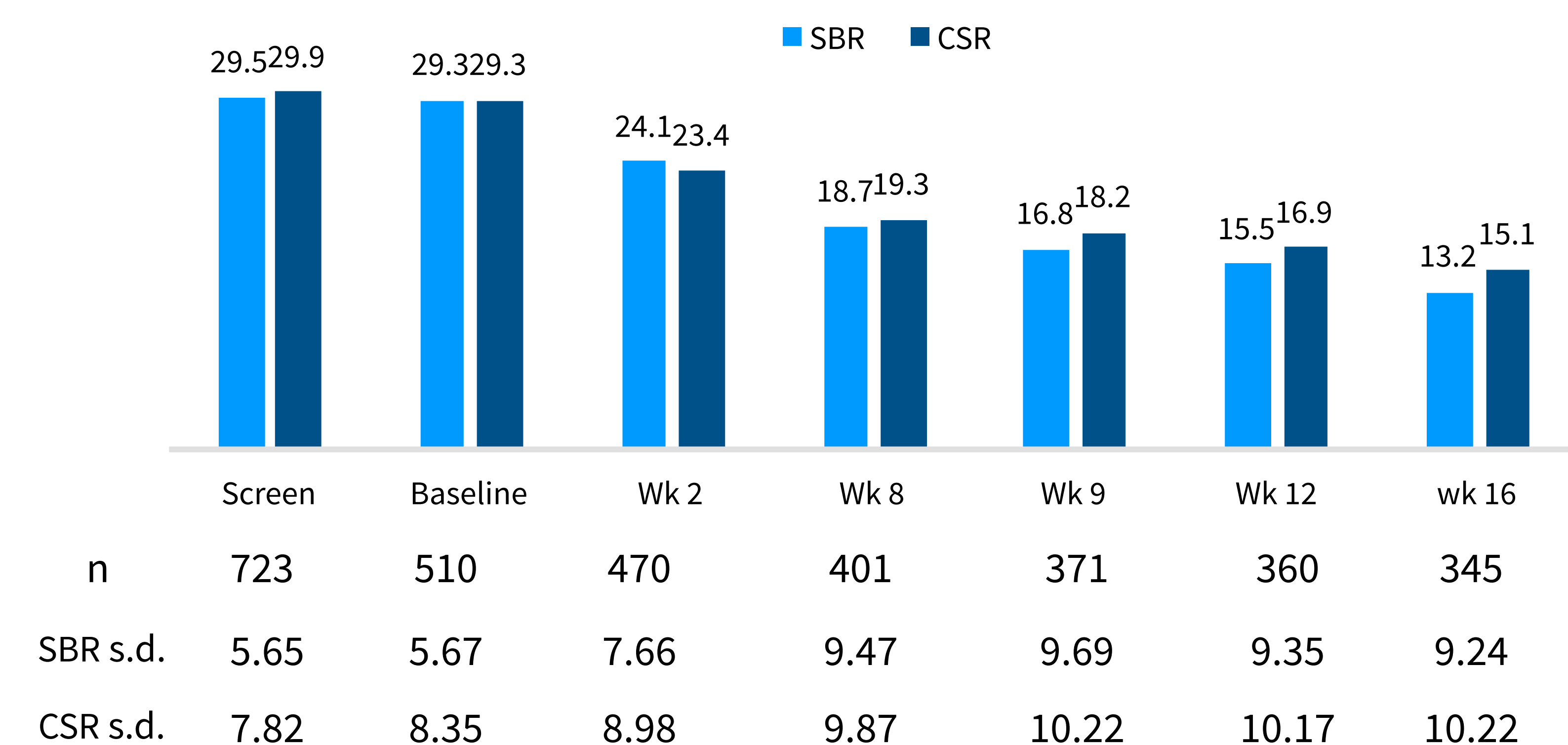
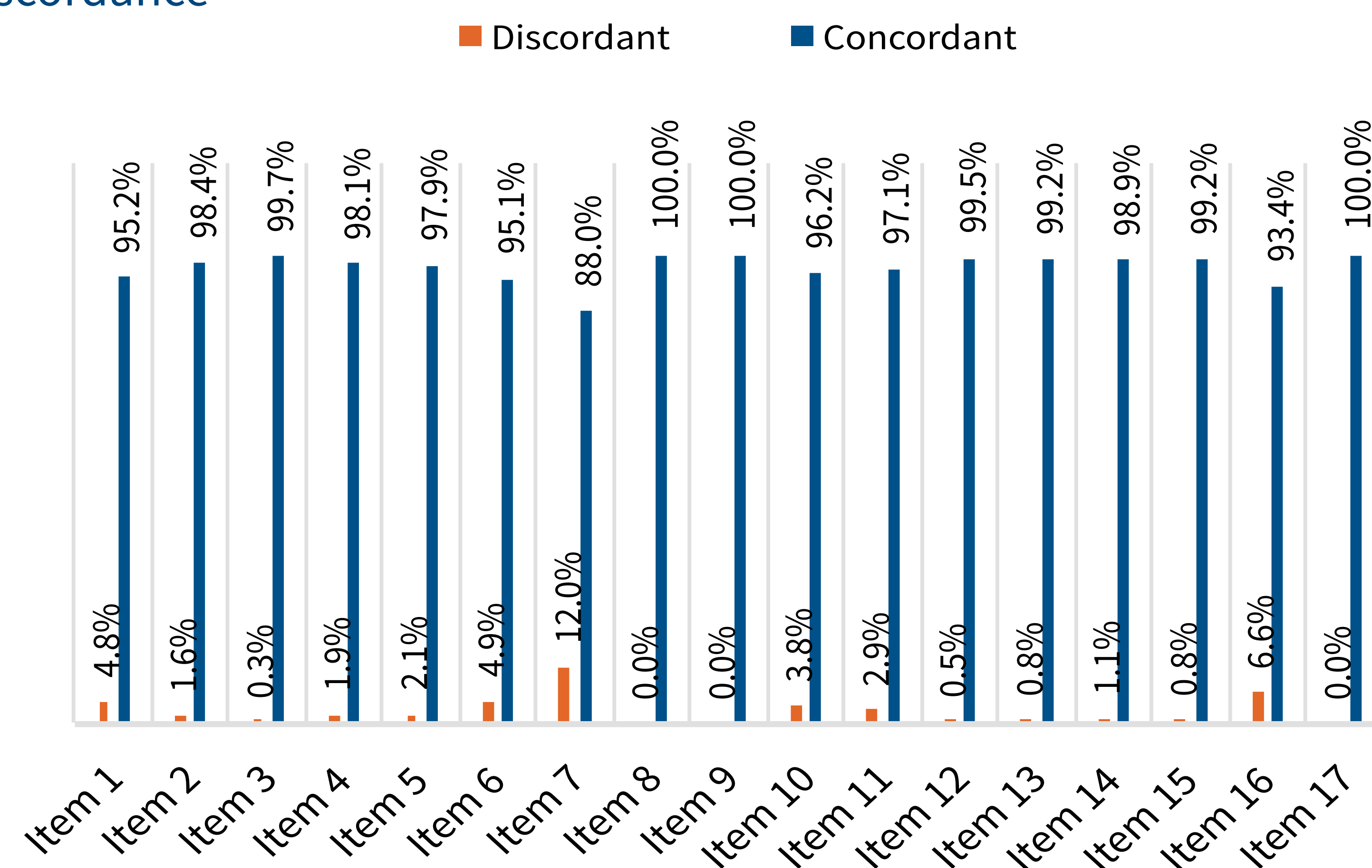


Figure 2: Item level Concordance (agreement SBR = CSR \pm 1) and Discordance



CONCLUSIONS

- SBR and CSR produced comparable mean total scores across all time points examined. In this actual RCT, item concordance for independent ratings was within the range required for rater certification. Changes in ICC and p-values observed after randomization may reflect subject practice effects, changes in variance over the course of the study or alteration in rater or respondent behavior after determination of eligibility.
- Computer administered scales may offer important advantages not because a CSR is a better than the average site-based rater, but because the computer is consistent, fast, and frugal. By simulating the judgment of a human rater the CSR offers an alternative to reliance on self-report measures.

REFERENCES

- Muller MJ, Dragicevic A. Standardized rater training for the Hamilton Depression Rating Scale (HAMD-17) in psychiatric novices. *Journal of affective disorders* 2003;77:65-9.
- Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Archives of general psychiatry* 1988;45:742-7.
- Klein DF, Thase ME, Endicott J, et al. Improving clinical trials: American Society of Clinical Psychopharmacology recommendations. *Archives of general psychiatry* 2002;59:272-8.