Title

The Rater Applied Performance Scale: Evaluating Clinical Interview Skill via Audio Recordings of MADRS Assessments in a Clinical Drug Trial

Authors

Engelhardt, N[1], Yavorsky, C[1], McNamara, C[1], Wolanski, K, Burger, F[1], Di Clemente, G[1].

[1] Cronos Clinical Consulting Services, Inc.

The Methodological Question Being Addressed

The quality of clinical interviews in CNS drug trials is frequently overlooked in favor of establishing interrater reliability with passive scoring tasks such as rating patients from a video recorded interview. What is the level of clinical interview skill of raters in a global multicenter clinical drug trial, and does clinical interview skill, as measured by the Rater Applied Performance Scale, bear any relationship to scoring accuracy?

Introduction

Audio recording monitoring of primary outcome clinical interviews has been applied to multi-center psychiatry trials as a way to monitor the quality of outcome data. The Rater Applied Performance Scale (RAPS) (Lipsitz et al, 2004) was developed to provide a more systematic and objective assessment of applied rater performance. The RAPS evaluates the clinical interview skill of raters as well as how reliably raters apply scoring criteria. It has been used in structured interview guide development, training, and active monitoring of rater performance in a clinical trial.

Methods

The RAPS measures six domains of clinical assessment: Adherence to scale administration guidelines, Follow-up questioning, Clarification of ambiguous responses, Neutrality, Rapport, and scoring Accuracy. Each domain is rated as Excellent, Good, Fair, and Unsatisfactory. Using domain ratings and pass/fail criteria developed for this analysis, we analyzed individual domain and total RAPS scores of 585 audio recorded interviews of the Structured Interview Guide for the Montgomery-Åsberg Depression Rating Scale (SIGMA) in a randomized, double blind placebo-controlled depression trial. Audio recorded interviews were conducted by independent blinded raters over the telephone.

Results

35.4% of raters failed the RAPS evaluation; the majority (64.6%) passed. The MADRS was judged to be scored accurately by 72.6% of the raters. All RAPS domains had a significant relationship to outcome (Pass/Fail) with the exception of Rapport (Chi-Square score of 4.542, p = .604). Virtually all of the raters (91.5%) received a score of excellent for this domain. Follow-up emerged as the domain that raters had the most difficulty executing: 28% received a Fair or Unsatisfactory rating. Follow-up was associated with scoring accuracy. There was a 72% probability that raters who were Good or Excellent in Follow Up met criteria for passing

Accuracy (n = 512). However, for those raters who received Fair or Unsatisfactory in Follow Up (n = 72), the probability of passing Accuracy was 44% and not substantially different than the probability of failing Accuracy (56%).

Conclusions

The majority of independent raters in this sample demonstrated adequate clinical interview skill and scoring accuracy. Independent raters, in addition to being blind to study visit and protocol, received extensive training and regular, ongoing calibration, which may differentiate their rating performance from site raters. Raters who engage in appropriate use of follow-up questions to elicit sufficient information tend to score more accurately than raters who do not. Rapport, thought to be critical in mitigating placebo response, was not related to RAPS outcome.

Lipsitz, J, Kobak, KA, Feiger, A, Sikich, D, Moroz, G, Engelhardt, N. The Rater Applied Performance Scale (RAPS): Development and Reliability. Psychiatry Research, 2004; 127: 147-155.