

Risk-Based Error Prevention Using Virgil Flags in PANSS Administration and Scoring

Negash, S¹, Stein, L¹, Prochnik, E¹, Webber-Lind, B¹, Detke, M^{1,2}, Williams, JBW^{1,3}

¹MedAvante, Inc, ²Indiana University of Medicine, ³College of Physicians and Surgeons, Columbia University

The Methodological Question Being Addressed

The Positive and Negative Syndrome Scale (PANSS) is a complex scale prone to administration and scoring errors that contribute to poor interrater reliability. A platform for electronic clinical outcome assessments (eCOA) provides triggers with real-time clinical guidance that alert the rater to scoring inconsistencies. The extent to which this type of targeted, risk-based approach identifies scoring errors in clinical trials, however, is less studied.

Introduction

Sponsors of clinical trials seeking ways to improve oversight efficiency and ensure the quality of study data are adopting technology-based solutions such as centralized risk-based monitoring (RBM). Another aspect of data quality is improving accuracy in the scoring of rating scales, a goal which also lends itself to a risk-based technology solution capable of identifying scoring errors in efficacy measures before the data are submitted. The PANSS is a complex scale prone to administration and scoring errors that can contribute to poor interrater reliability and inaccurate clinical trial results. The Virgil Investigative Study Platform, an eCOA platform that collects electronic source (eSource) data, is designed to standardize administration of diagnostic and outcome measurements and improve data quality by providing raters with clinical guidance in real time, including “flags” to alert the rater to scoring inconsistencies while administering the PANSS. In the present study, we examined the extent to which occurrences of these flags are associated with error rates in PANSS scoring in previous schizophrenia trials.

Methods

768 paper-based assessments aggregated from eight randomized, double blind, placebo-controlled schizophrenia trials were reviewed by the same central cohort of blinded clinicians who identified scoring errors by reviewing audio recordings and worksheets from each assessment. The scoring errors were then compared against the Virgil flags to determine how often a flag would have been triggered in assessments that had two or more errors. 18 Virgil flags, developed to identify the potential risk of PANSS scoring errors, range from within-visit scoring inconsistencies (e.g., a difference of more than two points between related items) to between-visit alerts (e.g., same response on all items from previous visit). Each flag was examined to determine its association with scoring errors.

Results

565 assessments (74%) had two or more scoring errors, indicating that scrutiny of PANSS ratings scale administration is essential. The flags that occurred with higher frequency in these problematic ratings included: depressed or low mood rated two or higher without verbal confirmation from the subject (71%); inconsistencies between distrustfulness and active social avoidance (33%); inconsistencies between lack of insight and delusions (17%); and between-visit total score change of greater than 25% (12%). Assessments with zero or one error triggered significantly fewer flags compared to those with two or more errors.

Conclusions

Risk-based analysis is useful in identifying potential scoring errors and technology can assist in preventing problematic ratings by directing the rater's focus to the most critical data elements that need attention. The Virgil eCOA platform with real-time clinical guidance, scoring anchors, and item descriptors minimizes scoring inconsistencies to reduce error variance and improve signal detection.