



What can data monitoring tell us about where to focus efforts to remediate problematic scoring of the Positive and Negative Syndrome scale?

Nina Engelhardt, PhD*; Matthew C. Masotti*, Kristy Wolanski, MA, LPC*; Francisco Burger, Psy.M*; Christian Yavorsky, PhD*; Guillermo Di Clemente, MSW, PhD*

*all authors are employees of Cronos CCS Inc., Lambertville, NJ, USA

Introduction

Many pharmaceutical companies support the use of prospective data monitoring (DM) of outcome measures in CNS trials. Such programs identify potential inconsistencies and errors in scoring during the course of a trial that are believed to affect the reliability and accuracy of ratings. A DM program is deemed successful if it recognizes and corrects scoring errors, reducing the frequency of errors over time and the need for remediation. Despite widespread use of DM, relatively little has been published that characterizes the nature, frequency and distribution of errors during the course of a trial. A recent study (Daniel & Kott, 2014) characterized risks to PANSS data integrity as “1) Large between-visit PANSS changes; 2) erratic PANSS changes; and 3) 100% identical PANSS scores from visit to visit.” Risk-based DM is a recognized, FDA sanctioned sponsor responsibility (FDA, 2013). There is a need for more information concerning the variety and frequency of risks to data integrity, as well as the impact of efforts to reduce risk. This research evaluates how the frequency and distribution of categories of scoring inconsistencies in the PANSS identify where to focus remediation efforts.

Methods

DM data was sampled from a global schizophrenia trial consisting of PANSS assessments. DM involved the utilization of proprietary algorithms to identify and monitor inconsistencies in measurement, expert clinical review, and remediation of rater performance. Some scoring inconsistencies reflected sufficient risk to data quality to necessitate a phone call with the rater to determine scoring rationale and to remediate where appropriate.

To obtain a manageable sample dataset of assessments requiring remediation, the 26 countries in the study were ranked by the number of PANSS assessments (22,129 total). The approximate middle third of the countries were selected, which comprised 4,439 assessments, 227 of which required remediation. The 227 assessments represented 10 countries, 68 raters, and 150 remediation calls. A remediation call could address more than one scoring issue.

Categories of scoring inconsistency or “Reasons for Call” were chosen based on their ability to predict rater error, i.e., once a scoring inconsistency is identified, errors are not confirmed until a case discussion occurs between the rater and a clinical specialist. The categories were Low Variability, Clinically Improbable Score Pattern, Inconsistent Item Relationships, Large Score Change and Other.

- **Low Variability:** 90% or greater of items with identical scores across visits.
- **Clinically Improbable Score Pattern:** tendency to consistently assign the same score to several or all items within the PANSS.
- **Inconsistent Item Relationships:** lack of association among items within a scale or across scales that measure similar constructs, e.g., a PANSS rating of 6 (Severe) on P6 (Suspiciousness/ Persecution) and a rating of 2 (Minimal) on P1 (Delusions).
- **Large Change Score:** $\geq 40\%$ decrease from screening in PANSS total score within 26 weeks or at least a 30% increase from screening in the first 10 weeks.

Results

The most frequent scoring problem was Inconsistent Item Relationships, which represented 38% of the sample. Low Variability, Clinically Improbable Score Pattern, Large Change Score, and Other were 31%, 20%, 6%, and 7%, respectively.

Figure 1 shows, on average, the number of calls made to a rater demonstrating Large Change Score was about half the number of calls made to raters demonstrating Low Variability, Inconsistent Item Relationships, and Clinically Improbable Score Pattern, each of which were associated with approximately the same number of calls.

Figure 2 takes into account the average number of days from PANSS assessment date that a scoring problem persisted. Inconsistent Item Relationships and Low Variability each persisted about 20% longer in days than both Clinically Improbable Score Pattern and Large Change Score. The distribution of scoring inconsistencies was different among the 10 countries. Figure 3 represents, by country, the number of raters who demonstrated a specific scoring problem. Eight out of 11 raters in Argentina and 11 out of 14 raters in Spain demonstrated Inconsistent Item Relationships. In addition, PANSS assessments characterized by Low Variability were a common feature in many countries. Given the relatively small size of the samples in each country, any generalizations regarding cultural differences in PANSS scoring should not be made.

Figure 1. Average number of calls per rater across categories of scoring inconsistency

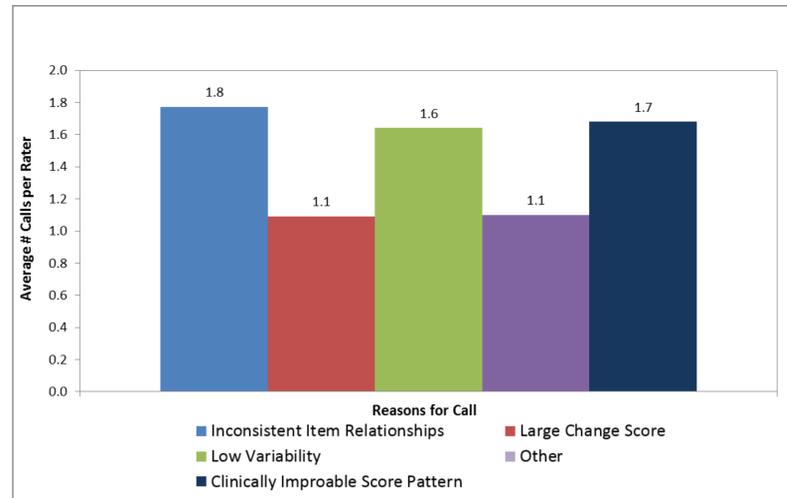


Figure 2: Average number of days a scoring problem persisted since PANSS assessment date

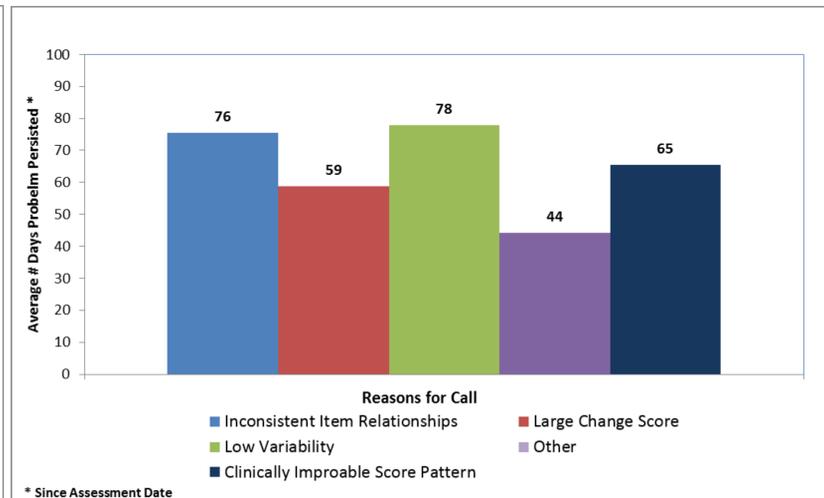
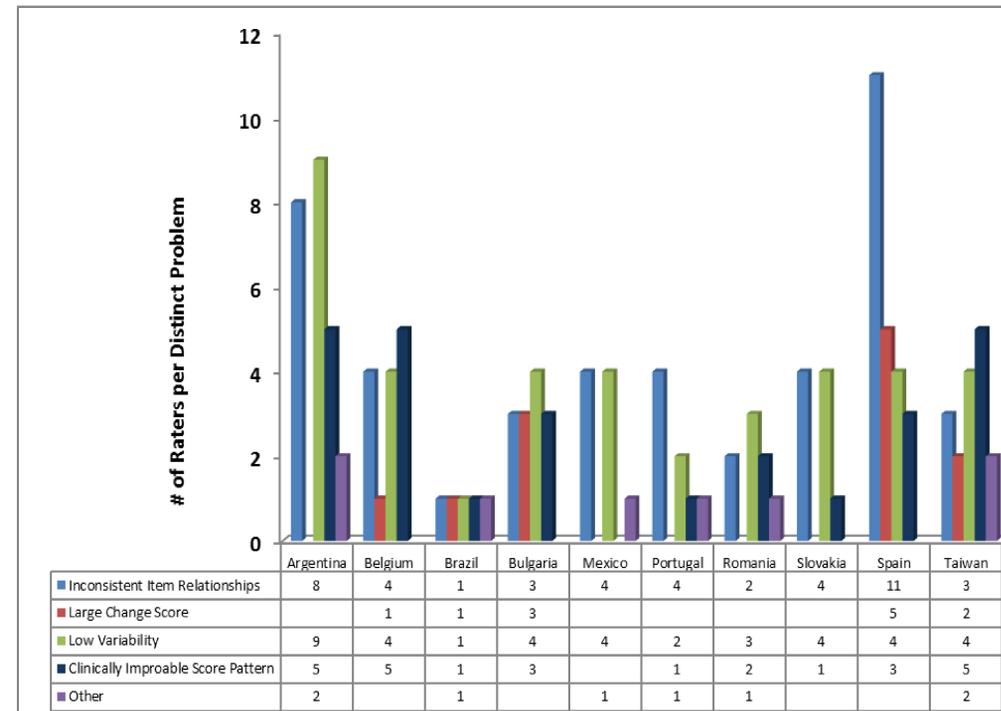


Figure 3: Number of raters per country demonstrating specific scoring inconsistencies



Conclusion

The distribution of categories of PANSS scoring inconsistencies illustrate that certain problematic scoring behaviors occur more frequently than others and may also be less amenable to change. Inconsistent Item Relationships was the most frequent scoring inconsistency and the least susceptible to change, as represented by number of calls per rater. Large Change Score, a relatively infrequent problem, was associated with the fewest number of calls per rater, suggesting this problem was more easily remedied.

More complex problems, such as Inconsistent Item Relationships and Clinically Improbable Score Pattern may require adjunct rater tools and more frequent calls. Comparison of problematic scoring categories from other researchers' PANSS data (Khan, A, et al, 2013; ISCTM Working Group, 2014) may enable refinement and standardization of these categories. Future efforts to characterize the frequency and distribution of scoring inconsistencies in larger samples may provide a rationale for targeting certain rater behaviors both prior to and during study implementation.

References

- Daniel, DG, Kott, A: Risk based data quality monitoring utilizing data analytics and recorded PANSS interviews in Global schizophrenia trials. Poster presented at the 10th Anniversary International Society of Clinical Trial Methodology (ISCTM) Meeting, Washington, DC, 18-20 February 2014.
- U.S. DHHS, FDA, CDER, CBER, CDRH, OGCP, ORA: Guidance for Industry: Oversight of clinical investigations--a risk-based approach to monitoring. OMB Control # 0910-0733, August 2013.
- Khan, A, Yavorsky, WC, Di Clemente, G, Rothman, B, Opler, M, DeFries, A, Jovic, S. Assessing the sources of unreliability (rater, subject, time point) in a failed clinical trial using items of the PANSS. *J Clin Psychopharmacol* 2013; 33:109-117.
- Use of algorithms/flags to identify clinical inconsistency. 10th Anniversary International Society of Clinical Trial Methodology (ISCTM) Working Group, Washington, DC, 18-20 February 2014.

*Corresponding Author: Nina Engelhardt, nina.engelhardt@cronosccs.com, Disclosure: The authors report no conflicts of interest for this work.