

# A NEW PLATFORM TO IMPROVE QUALITY OF PANSS AND MADRS ADMINISTRATION

Williams JBW<sup>1,2</sup>, Echevarria B<sup>1</sup>, Osman D<sup>1</sup>, Garzio LM<sup>1</sup>, Detke MJ<sup>1,3,4</sup>

<sup>1</sup> MedAvante, Inc. <sup>2</sup> Dept. of Psychiatry, Columbia University College of Physicians and Surgeons, <sup>3</sup> Detke Biopharma Consulting LLC, <sup>4</sup> Indiana University School of Medicine

## INTRODUCTION

Strategies are needed to improve quality of rater administration of many subjective psychiatric instruments.

- **The Challenge:** Two commonly used measures of psychopathology, the Structured Clinical Interview for Positive and Negative Syndrome Scale (SCI-PANSS) and the Structured Interview Guide for the Montgomery-Asberg Depression Rating Scale (SIGMA) are complex scales to administer and often difficult to score.
  - SCI-PANSS raters are required to score 30 different items across three major subscales.
  - Each PANSS item has unique description, rating guidelines, scoring anchors and scoring conventions and some are not clearly defined in the PANSS manual.
  - SIGMA raters are required to consider six potential levels of intensity and frequency for each of the ten Montgomery-Asberg (MADRS) items.
  - Every MADRS item has its own unique description, rating guidelines, scoring anchors and conventions.
- **The Solution:** Real-time clinical guidance and scoring assistance provided to raters while administering the SCI-PANSS and SIGMA has been developed as part of a new electronic source (eSource) data capture and monitoring platform with the potential to decrease errors and increase interrater reliability.
  - We tested the ability of the eSource platform's automated scoring alerts to improve quality of rater administration of the SCI-PANSS and the SIGMA with recently collected clinical trial data.

## METHODS

- Called upon extensive training and administration experience (>25,000 SCI-PANSS and >39,000 SIGMA central assessments completed) to identify the most common administration errors.
- Selected one type of electronic alert - entry of incompatible item scores - that the eSource platform would automatically provide to raters in real-time.
- Identified eleven pairs of incompatible PANSS item scores, either within the same subscale or between different subscales.
  - For example, a positive score on Tension (G4) requires a verbal attestation to the presence of Anxiety (G2). This convention is noted in the PANSS manual and usually covered in training but frequently forgotten by raters.
- Identified two additional pairs of incompatible item scores across the ten MADRS items.
  - For example, a higher rating ( $\geq 4$ ) on Reported Sadness would suggest the presence of pessimistic thinking, but if raters score Pessimism as a zero they may not have considered any discouragement or pessimism noted in their Reported Sadness score.
- These thirteen potential errors (each of which would generate an "alert" when entered into the eSource platform) were used to analyze a subset of PANSS assessments (n=288, 79 raters) and SIGMA assessments (n=1200, 111 raters) recorded in recent clinical trials.
  - Raters' scores were analyzed to determine how many items would have triggered at least one alert had the real-time interventions been available as they scored the assessments.

## CONCLUSION

- While the PANSS and the MADRS are significantly different instruments in terms of number of items and construct overlap, scoring accuracy of both may improve if incompatible item scores are flagged for raters in real time.
  - In addition to alerts to reduce scoring errors, an eSource platform with multi-level clinical guidance (e.g., links to scoring anchors, item descriptions and study-specific rating guidelines) during each interview can improve administration, increase interrater reliability and thus strengthen the ability to detect a positive signal.
- Not all scoring alerts indicate a scoring error, but they should prompt raters to reconsider conflicting ratings in light of scoring conventions and potential clinical inconsistencies.
  - Error rates may not replicate in other studies since assessments were selected for review based on different criteria in each study.

## RESULTS

- 135 (47 percent) of SCI-PANSS interviews would have triggered at least one alert. (Figure 1)
  - Because some assessments had more than one alert, a total of 171 individual alerts would have been triggered.
  - 58 (73 percent) of the SCI-PANSS raters would have been alerted to at least one potential error. One rater accounted for 11 percent of scoring alerts and the remaining alerts were distributed among the other raters.
- 77 (6.4 percent) of the SIGMA interviews would have triggered at least one alert.
  - 44 (40 percent) of the SIGMA raters would have been alerted to a potential scoring issue.

FIGURE 1. Number and percentage of alerts per SCI-PANSS assessment

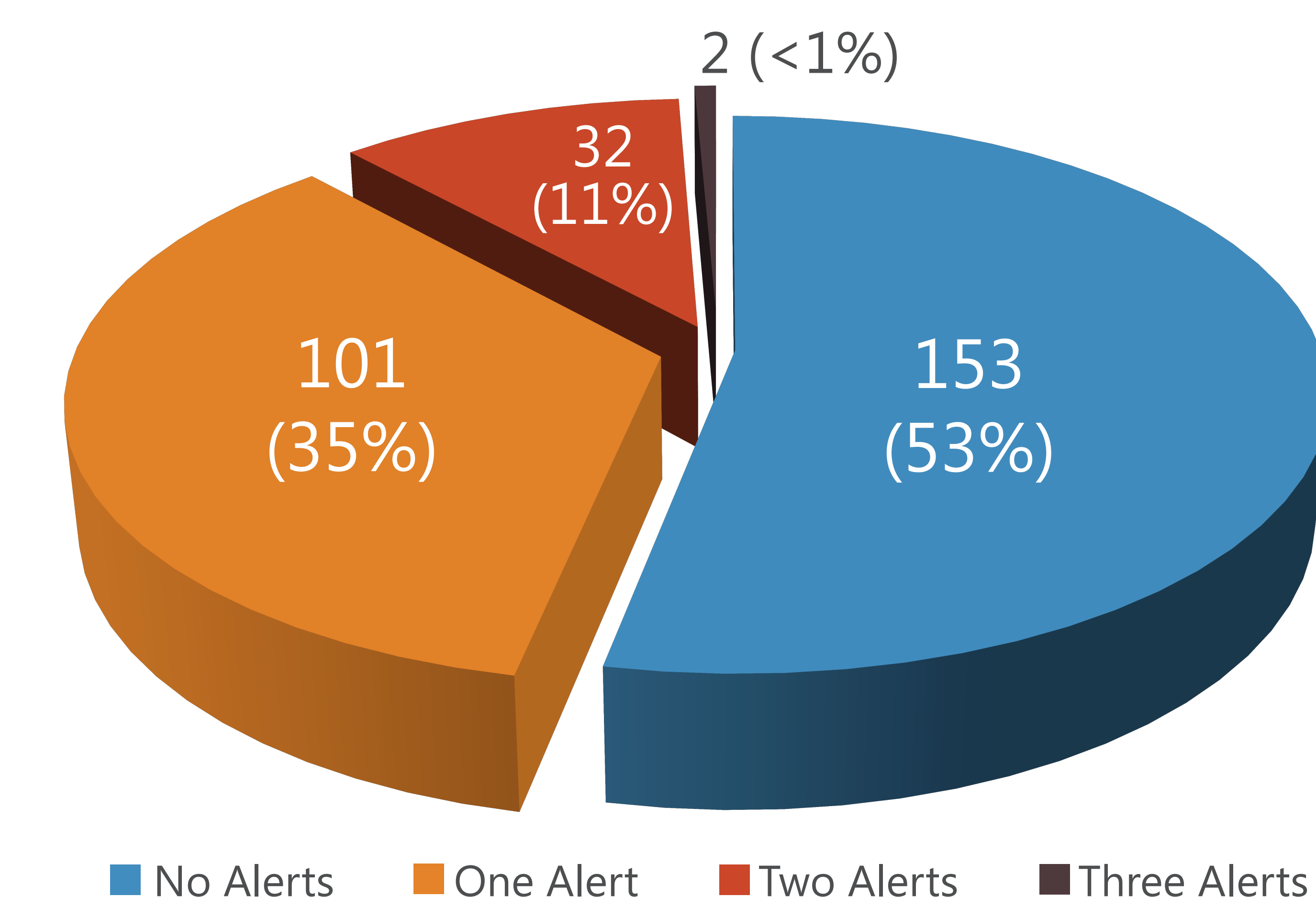
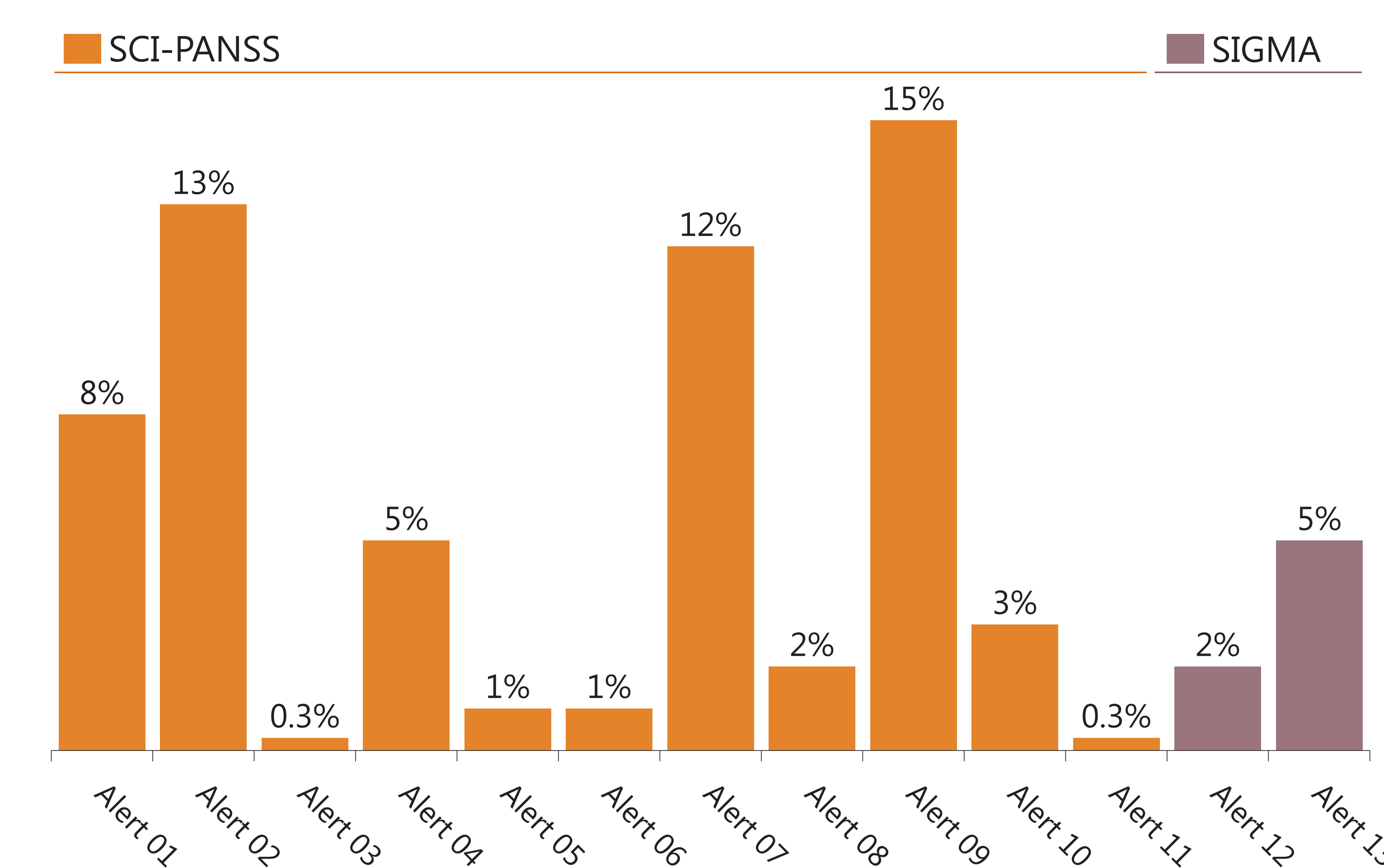


FIGURE 2. Percentage of SCI-PANSS and SIGMA assessments flagged per scoring alert



- The percentage of assessments that generated alerts for each of the incompatible scores varied according to the alert. (Figure 2)
  - Three of the SCI-PANSS alerts would have been triggered in more than twelve percent of the assessments.
  - Five of the SCI-PANSS alerts would have been triggered in more than five percent of the assessments.
  - One of the SIGMA alerts would have been triggered in five percent of the assessments.

