How Do US Sites Compare with Rest of World in PANSS Interview and Ratings Quality?

INTRODUCTION
As drug-placebo differences have diminished over time clinical trials sites in the United States have come under increasing scrutiny (1,2). We compare the quality of interview and ratings procedures delivered by North American (NA) vs. rest of world (ROW) sites using pooled results of video/audio surveillance measures from ten international schizophrenia clinical trials.

METHOD
A proprietary video/audio recording system was utilized to record PANSS rating procedures in ten schizophrenia clinical trials conducted in North America, Europe, South America and Asia. Sites uploaded video or audio recorded ratings assessments for review by calibrated external reviewers. External reviewers provided feedback on ratings quality on an ongoing basis to the site and sponsor. Prior to study initiation, raters were trained at investigator’s meetings by highly interactive procedures, including slide presentations, rating of videotaped patient interviews, and in some cases, interview and rating of live actors trained to portray schizophrenia symptoms. Instruments used to assess ratings and interview quality included the Research Interview Assessment Scale (RISA) (3) and a new tool, the Rater Quality Questionnaire (RQQ) (4).

The RISA is a 16 item scale representing 4 domains of interview quality. Higher scores represent better quality interviews. The RQQ is a global measure of interview and ratings quality consisting of two domains addressing: 1) the overall quality of the patient and/or informant material collected; and 2) the overall quality of the ratings or diagnostic evaluation (including proper application of the rules and anchor points of the rating scale or structured interview) for the current subject visit. Each domain is evaluated on a Likert-like scale ranging from 1-3. A score of 1 represents no deficiencies in information gathering and/or scoring. A score of 2 represents deficiencies that have a minor impact on interview quality or scoring. A score of 3 reflects serious deficiencies in either collecting the information or rating the symptoms observed.

RESULTS

SUMMARY OF RESULTS I: 1901 videotaped PANSS administrations at sites were graded for interview quality by external reviewers using the RISA. Mean total RISA scores of the North American (NA) raters (27.9 +/- 3.00) were modestly, but statistically significantly lower (indicating lower quality) than rest of world raters (ROW) (28.3 +/- 2.54) (t(1899)=2.94, n=1901, p<0.01).

Figure 1: Quality of Interviews as Judged by the RISA Scale: North America vs. Rest Of World

SUMMARY OF RESULTS II: 1616 videotaped PANSS administrations at sites were rated for interview quality by external reviewers using the RQQ global interview quality axis. Mean total RQQ interview subscale scores of the North American (NA) raters (1.27 +/- 0.49) were modestly, but statistically significantly lower (indicating higher quality) than rest of world raters (ROW) (1.34 +/- 0.50) (t(1614)=2.54, n=1616, p<0.05).

Figure 2: Quality of Interviews as Judged by the RQQ Scale: vs. Rest Of World

SUMMARY OF RESULTS III: 1616 videotaped PANSS administrations at sites were rated for proper application of the anchor points and other instructions of the measurement tool by external reviewers using the RQQ global ratings quality axis. Mean total RQQ scores of the North American (NA) raters (1.27 +/- 0.47) were not statistically significantly different than rest of world raters (ROW) (1.30 +/- 0.49) (t(1614)=1.26, n=1616, p=NS).

Figure 3: Quality of Ratings as Judged by the RQQ Scale: North America vs. Rest Of World

DISCUSSION
North American raters scored as well or better than their ROW counterparts on the RQQ which evaluates the quality of information collected by interview and adherence to rating scale rules. On the RISA total score which evaluates a broad range of interview behaviors (eg, ability to clarify, adherence to structured interview guidelines, handle questions that may induce placebo response) NA raters scored modestly worse than their ROW counterparts. This report is preliminary as data continues to be collected in ongoing clinical trials.

REFERENCES

Table I: Measurement Scale Challenges in Global Schizophrenia Trials

- Long, complex rating scales
- Broad, diverse range of symptoms
- Requirements for caregiver input
- Visual, auditory and report data required
- Self report not be reliable
- Culturally diverse interpretation of symptoms

Table II: Rater Quality Questionnaire

<table>
<thead>
<tr>
<th>Domain</th>
<th>Anchor Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the overall quality of the patient and/or informant material collected for this subject visit?</td>
<td>1. The patient and/or informant material required to complete the rating scale is clearly documented. 2. The patient and/or informant material required to complete the rating scale is deficient in some cases but overall is adequate to rate the patient. 3. The patient and/or informant material required to complete the rating scale is inadequate to rate the patient.</td>
</tr>
<tr>
<td>What is the overall quality of the site raters ratings for this subject visit?</td>
<td>1. The ratings reflect proper application of the anchor points and other instructions of the measurement tool. 2. The ratings are too high. 3. The ratings are too low. 4. The ratings reflect improper application of the anchor points and other instructions of the measurement tool. 5. Overall, the ratings reflect improper application of the anchor points and other instructions of the measurement tool.</td>
</tr>
</tbody>
</table>

Daniel D1; Kott A2
1Bracket, McLean, VA, USA; 2Bracket, Prague, CZ