Evidence That Sites Can Conduct High Quality Interviews and Ratings in Global Clinical Trials

David G. Daniel, MD1, Alan Kott, MD2 and Amir Kalali, MD3
1Bracket, McLean, VA, USA; 2Bracket, Prague, CZ; and 3Quintiles, San Diego, CA, USA

INTRODUCTION

As drug-placebo differences have diminished over time in drug-placebo differences in schizophrenia clinical trials, the quality of interview and rating procedures performed by sites have come under increasing scrutiny (1-4). We describe the pooled results of 1442 videotaped PANSS administrations at sites graded for interview quality by external reviewers using the RISA (1.35±.51 and 1.31±.50, respectively) and ratings reflected improper application of the anchor points and other instructions of the measurement tool; 1.7% (n=20) the ratings reflected improper application of the anchor points and other instructions of the measurement tool.

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 Bracket. Prior to study initiation, raters were trained at investigator 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) (5) and a new tool, the Rater Quality Questionnaire (RQQ) (6). The RQA is a 16 item scale representing 4 domains of interview quality. Higher scores represent better quality interviews.

The RQA 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 for the current subject visit; 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 (1-3). A score of 1 represents no deficiencies in information gathering and/or scoring. A score of 2 represents deficiencies that have a minimal impact on the scoring. A score of 3 reflects serious deficiencies in either collecting the information or rating the symptoms observed.

RESULTS

SUMMARY OF RESULTS I: 1442 videotaped PANSS administrations at sites were graded for interview quality by external reviewers (RQA Score 28-30), 28.9% as acceptable (RISA score 24-27) and only 6.5% as poor or unacceptable.

SUMMARY OF RESULTS II: For the RQA analysis, the initial pooled dataset consists of 1187 evaluated interviews. The mean RQA scores for sufficiency of the data collected by interview and competency in use of the rating scales were 1.35±.51 and 1.31±.50, respectively.

SUMMARY OF RESULTS III: In 67.3% (n=799) of visits the patient and/or informant material required to rate the patient was classified as fully adequate; 30.8% (n=366) as acceptable (RISA score 24-27) and only 2.9% (n=36) as inadequate to rate the patient.

SUMMARY OF RESULTS IV: In 71.1% (n=844) of visits the ratings were felt to reflect proper application of the anchor points and other instructions of the measurement tool; in 27.2% the ratings generally, but not uniformly reflected proper application of the anchor points and other instructions of the measurement tool; and in 1.7% (n=20) the ratings reflected improper application of the anchor points and other instructions of the rating tool.

SUMMARY OF RESULTS V: 97.5% of the interviews were graded as adequate with no more than minimal deficiencies in any area. Only 1.4% of interviews had serious deficiencies in one of the domains and only 1.1% of interviews had serious deficiencies in both domains. Most (60.2%) interviews were considered as fully proficient in both information gathering and adherence to PANSS scale instructions. The remainder had at least minor deficiencies. 7.2% of interviews had good quality of information gathering but had minor difficulties in scoring this information; 10.8% of interview had minor deficiencies in information gathering but the information was scored using proper technique. 19.5% of interviews had minor deficiencies in both information gathering and scoring.

DISCUSSION

The initial data indicate that in global clinical trial settings with surveillance of ratings and ongoing feedback to investigators the quality of interview data and proficiency of ratings were judged to be adequate or better by external reviewers in the large majority of cases. These results are preliminary and additional data will be reported from ongoing and recently completed studies. This study’s findings differed from those of Jeglic et al (2007) and Engelhardt et al (2006) in which the majority of interviews were judged to be deficient. We speculate that rapid feedback to the raters from the external reviewers may have been useful in maintaining interview and ratings quality in the current studies.

REFERENCES

4) Kobak KA; Inaccuracy in Clinical Trials: effects and methods to control inaccuracy. Curr Alz Res 2010 Nov; 7 (7) :637-641
6) Daniel GD and Holt A; Initial Findings of the Rater Quality Questionnaire (RQA): A New Tool for Evaluating Quality of Ratings in Global Clinical Trials, Poster to be presented at the 25th ECNP Congress, October 13-17, 2012, Vienna, Austria.

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 rater 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 generally, but not uniformly reflect proper application of the anchor points and other instructions of the measurement tool. 3. Overall, the ratings reflect improper application of the anchor points and other instructions of the measurement tool.</td>
</tr>
</tbody>
</table>

Table III: Summary of RQQ Findings

<table>
<thead>
<tr>
<th>Rater Quality Questionnaire (RQQ)</th>
<th>Adherence to Scale Instructions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Patient and/or Informant Collected</td>
<td>128</td>
<td>736</td>
</tr>
<tr>
<td>(0.17%)(0.59%)</td>
<td>0.17%</td>
<td>(0.59%)</td>
</tr>
<tr>
<td>Total</td>
<td>844</td>
<td>323</td>
</tr>
<tr>
<td>71.11%</td>
<td>27.21%</td>
<td>1.68%</td>
</tr>
</tbody>
</table>