ABSTRACT

Determining the reliability of clinical rating scales across the range of circumstances of their proposed use is key to establishing the acceptability of any measure proposed as a drug development tool. Since most reports establishing scale and rater reliability derive their data from an exercise conducted apart from an actual clinical trial, there is a need to establish the performance characteristics of scales within the context of their intended use. Well established scales such as the Montgomery Asberg Depression Rating Scale (MADRS) may provide a useful benchmark against which to judge the merit of other proposed measures.

METHODS

Three multicenter placebo controlled double blind studies were identified in which the MADRS was administered independently by site based raters (MADRS_SBR) and by a computer (MADRS_COMP) as part of a quality management program. Internal scale consistency was assessed using Cronbach’s alpha calculated for the MADRS_SBR and MADRS_COMP at baseline, the first post randomization visit, study endpoint, other visits, and all visits. All ratings were made in the subject’s native language. Comparisons of the MADRS_SBR and MADRS_COMP were made at overall and at four study time points Baseline, First Post-randomization visit, Study Endpoint and all other visits. Correlations between the pairs were calculated to examine measurement reliability overall and individually for each MADRS item.

CONCLUSION

Guidance on the development of drug development tools makes clear the need for scale performance data. This analysis shows good reliability for the MADRS at all study visits after baseline. Close agreement between MADRS_COMP and MADRS_SBR across study visits and therapeutic indications during actual clinical trials supports the validity and reliability of computer administered MADRS in studies of Unipolar and Bipolar Depression. The MADRS_SBR and MADRS_COMP may be a useful benchmarks against which to judge the performance of other scales proposed for use in global multisite studies. The similar drop in performance observed for both the MADRS_SBR and MADRS_COMP at study baseline suggests that factors other than rater behavior may play an important role undercutting the reliability of baseline scores. Strengths and weaknesses of MADRS_SBR and MADRS_COMP should be considered in the selection of outcome measures.

REFERENCES


Gary Sachs, MD1; Michelle Arkow2; Dan DeBonis2

1Bracket, Lexington, MA; 2Bracket, Wayne, PA

BACKGROUND

Determining the reliability of clinical rating scales across the range of circumstances of its proposed use is key to establishing the acceptability of any measure proposed as a drug development tool. Since most reports establishing scale and rater reliability derive their data from an exercise conducted apart from an actual clinical trial, there is a need to establish the performance characteristics of scales within the context of their intended use. Well established scales such as the Montgomery Asberg Depression Rating Scale (MADRS) may provide a useful benchmark against which to judge the merit of other proposed measures.

METHODS

Three multicenter placebo controlled double blind studies were identified in which the MADRS was administered independently by site based raters (MADRS_SBR) and by a computer (MADRS_COMP) as part of a quality management program. Internal scale consistency was assessed using Cronbach’s alpha calculated for the MADRS_SBR and MADRS_COMP at baseline, the first post randomization visit, study endpoint, other visits, and all visits. All ratings were made in the subject’s native language. Comparisons of the MADRS_SBR and MADRS_COMP were made at overall and at four study time points Baseline, First Post-randomization visit, Study Endpoint and all other visits. Correlations between the pairs were calculated to examine measurement reliability overall and individually for each MADRS item.

At each time point, the variance observed by site based rater was compared to that observed by the computer and expressed as a percentage of the variance observed for the computer administered MADRS. (MADRS_SBR/MADRS_COMP x 100).

RESULTS

The sample included 7544 pairs of MADRS_SBR and MADRS_COMP ratings. The datasets had no subject identifiers other than the subject’s study identification number.

### TABLE 1: Mean Scores and Cronbach’s Alpha Across Study Visits

<table>
<thead>
<tr>
<th>Study Visit</th>
<th>MADRS_SBR Mean</th>
<th>MADRS_COMP Mean</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>30.4 (8.6)</td>
<td>23.4 (10.2)</td>
<td>0.657</td>
</tr>
<tr>
<td>First Post-randomization</td>
<td>23.4 (10.2)</td>
<td>15.0 (11.6)</td>
<td>0.829</td>
</tr>
<tr>
<td>Study Endpoint</td>
<td>18.1 (11.5)</td>
<td>18.1 (11.5)</td>
<td>0.842</td>
</tr>
<tr>
<td>Other Visits</td>
<td>20.3 (11.9)</td>
<td>20.3 (11.9)</td>
<td>0.868</td>
</tr>
<tr>
<td>All Visits</td>
<td>20.7 (11.6)</td>
<td>20.7 (11.6)</td>
<td>0.877</td>
</tr>
</tbody>
</table>

### GRAPH 1: Mean MADRS Score Across Study Visits

- Baseline
- Post-randomization
- First Post-randomization
- Study Endpoint

### GRAPH 2: Cronbach’s Alpha Across Study Visits

- Baseline
- Other visits
- Study Endpoint
- All visits

### GRAPH 3: SBR Variance as a Proportion of Computer Variance Across Study Visits

- Baseline
- First Post-randomization
- Other Visits
- Study Endpoint
- All Visits

- 40% SBR
- 60% Computer

Statistical analysis was funded by Bracket.

Presented at the 50th Anniversary Meeting of the American College of Neuropsychopharmacology (ACNP), December 4-8, 2011, Waikoloa Village, Hawaii.