A Retrospective Exploratory Study of the Variability of Radiologists' Measurements in a Selected Subgroup of Subjects Enrolled in a Clinical Trial

Robert W. Ford 1, Kathy Zhou Ph.D. 2, Robert R. Ford M.D. 3

1 Geisinger Health Care Systems, Danville, PA | 2 Department of Health Science Research, Geisinger Health System, Danville, PA | 3 Department of Health Science Research, Geisinger Health System, Danville, PA

Background Information

- Independent centralized review of imaging studies in an established means of validating data used in support of corporate “go/no go” decisions, regulatory approval, and post-marketing claims
- This study was performed to investigate and assign a concise value(s) to the inherent human variability in radiological measurements, despite best efforts to standardize assessment using Response Evaluation Criteria in Solid Tumors (RECIST)

Materials and Methods

- An experienced radiologist (34 years) and a research intern reviewed Digital Computed Tomographic (CT) scans from 31 subjects to identify 160 primary and metastatic tumors. Tumors were categorized into 4 categories.

<table>
<thead>
<tr>
<th>Tumor Category</th>
<th>Tumor Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Defined Edge/Radially Symmetric</td>
</tr>
<tr>
<td>2</td>
<td>Defined Edge/Peripheral Shape</td>
</tr>
<tr>
<td>3</td>
<td>Blurred Edge/Radially Symmetric</td>
</tr>
<tr>
<td>4</td>
<td>Blurred Edge/Peripheral Shape</td>
</tr>
</tbody>
</table>

- Category 1
- Category 2
- Category 3
- Category 4

Training The Radiologist Readers

- Each radiologist reader attended a training session so that the study was performed consistently
- There was a presentation covering an example of lesion measurement and the correct method of recording obtained data in the form provided
- Also included in this presentation were the software operation and lesion measurement guidelines

Lesion Measurement

- At the time of the read, each radiologist was provided with written directions for software operation and lesion measurement
- Fifteen radiologists independently measured each tumor
- The zoom function was allowed to be used freely while windows and level were held constant for each lesion
- The axial location of each lesion was specified by the radiologist by an image number
- The lesion to be measured was identified by a numbered annotation remaining the lesion that did not describe its geometric boundaries
- The lesions were presented to each reader in a categorically arbitrary sequence
- Each reader measured the longest dimension of each lesion in the axial plane using electronic calipers
- All scans were digital and thus required no pixel calibration (Calibration factors were encoded in the DICOM header)
- All readers used the same validated DICOM based software application

Statistical Analysis

- Statistical analysis was performed using mixed effects modeling to partition error according to its contributing factors
- Mixed Effects modeling allows for the analysis of heterogeneous data

Results

- Explanation of Terms
  - Total Error – Accounts for all the variance seen in the measurements of the lesions
  - Between Reader Error – The portion of the variance in the measurements that the radiologist readers are responsible for
  - Random Error – The portion of the total error from all other contributing sources

- Significant Data
  - The between reader error accounted for 4.14% percent of total errors for the length
  - The total error is smallest for category 1 and largest for category 4 as expected

<table>
<thead>
<tr>
<th></th>
<th>No of Tumors</th>
<th>Between Reader Error</th>
<th>Random Error</th>
<th>Total Error</th>
<th>Reader Error/Total Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>150</td>
<td>0.32 (0.14, 0.77)</td>
<td>7.51 (7.06, 8.01)</td>
<td>7.83</td>
<td>4.15%</td>
</tr>
<tr>
<td>Cat. 1</td>
<td>54</td>
<td>0.19 (0.08, 0.45)</td>
<td>1.25 (1.14, 1.39)</td>
<td>1.45</td>
<td>13.4%</td>
</tr>
<tr>
<td>Cat. 2</td>
<td>40</td>
<td>0.26 (0.08, 0.76)</td>
<td>4.67 (4.16, 5.24)</td>
<td>4.93</td>
<td>52.3%</td>
</tr>
<tr>
<td>Cat. 3</td>
<td>16</td>
<td>0.44 (0.14, 1.39)</td>
<td>3.92 (2.34, 4.75)</td>
<td>4.36</td>
<td>10%</td>
</tr>
<tr>
<td>Cat. 4</td>
<td>40</td>
<td>0.81 (0.25, 2.56)</td>
<td>19.98 (17.72, 22.47)</td>
<td>20.79</td>
<td>3.90%</td>
</tr>
</tbody>
</table>

Conclusions

- The (lower reader error) / total error suggests that the radiologist readers were only responsible for a small portion of the total error found in this measuring process
- Due to the fact that the intra-reader variability is usually lower than the inter-reader variability one can speculate that these 15 radiologists reading one lesion would be comparable to 1 radiologist reading the same lesion 15 times

Continuing Research

- The structure and implementation of the project facilitates continuing research
- Analyze intra-reader variability by repeating the process a later date
- Repeating the project will also help to strengthen conclusions made based on the current data
- Other functions in the software can be used as the variable
- The results database can now be used to qualify new radiologist readers and relate previously qualified readers