Imaging Charters and Reader Metrics in Independent Radiology Review

Michael O’Neal, MD
Chief Medical Officer
RadPharm Imaging Core Lab
Topics

- Charter Process
- Individual Reader Metrics
- Discordance Rates
- Site vs. Central Review
Imaging Charters
Charter Process

- TOC and Lexicon developed at previous meetings

- Helpful if Imaging Charter could be submitted with protocol for Registrational Oncology trials with imaging as endpoint
  - ICLs involved early (assist with standardization of imaging requirements and site selection)
  - FDA review and acceptance of Imaging Charter
Clinical Data Process

• Pre-specified what clinical data is required in charter

• From sponsor in usable format (per patient, per time point)

• Only monitored, cleaned data

• Controlled process for handling updates to clinical data (deletions and changes to previously reviewed data)

• Timeline impacts
Monitoring Reviewer Metrics

© 2010 Confidential and proprietary property of CoreLab Partners, Inc, all rights reserved.
Monitoring Reader Metrics
“Consensus Agreement or Adjudication Win/Loss Rates”

• How do we monitor reader performance in 2 reader/adjudicator studies?

• Determine percentage of the time the adjudicator accepts or rejects an outcome on a per reader basis when the case is adjudicated.

<table>
<thead>
<tr>
<th>READER</th>
<th>TOTAL READ</th>
<th>TOTAL ADJUDICATED</th>
<th>% ADJUDICATED</th>
<th># ACCEPT</th>
<th># REJECT</th>
<th>% ACCEPTED</th>
<th>% REJECTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>READER A</td>
<td>125</td>
<td>37</td>
<td>30%</td>
<td>23</td>
<td>14</td>
<td>62%</td>
<td>38%</td>
</tr>
<tr>
<td>READER B</td>
<td>2295</td>
<td>805</td>
<td>35%</td>
<td>371</td>
<td>434</td>
<td>46%</td>
<td>54%</td>
</tr>
<tr>
<td>READER C</td>
<td>549</td>
<td>229</td>
<td>42%</td>
<td>92</td>
<td>137</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>READER D</td>
<td>1197</td>
<td>457</td>
<td>38%</td>
<td>244</td>
<td>213</td>
<td>53%</td>
<td>47%</td>
</tr>
<tr>
<td>READER E</td>
<td>726</td>
<td>265</td>
<td>37%</td>
<td>152</td>
<td>113</td>
<td>57%</td>
<td>43%</td>
</tr>
<tr>
<td>READER F</td>
<td>1153</td>
<td>414</td>
<td>36%</td>
<td>233</td>
<td>181</td>
<td>56%</td>
<td>44%</td>
</tr>
<tr>
<td>READER G</td>
<td>734</td>
<td>294</td>
<td>40%</td>
<td>112</td>
<td>182</td>
<td>38%</td>
<td>62%</td>
</tr>
<tr>
<td>READER H</td>
<td>662</td>
<td>236</td>
<td>36%</td>
<td>109</td>
<td>127</td>
<td>46%</td>
<td>54%</td>
</tr>
<tr>
<td>READER I</td>
<td>1495</td>
<td>571</td>
<td>38%</td>
<td>328</td>
<td>243</td>
<td>57%</td>
<td>43%</td>
</tr>
<tr>
<td>READER J</td>
<td>724</td>
<td>318</td>
<td>44%</td>
<td>142</td>
<td>176</td>
<td>45%</td>
<td>55%</td>
</tr>
</tbody>
</table>

© 2010 Confidential and proprietary property of CoreLab Partners, Inc, all rights reserved.
Consensus Agreement Rates
29 Protocols, 3,944 Adjudications, 24 Readers

• Accept/Reject rates tend to fall along the 50% line
• Suggests that many adjudications are based on justifiable differences between readers
### Reasons for Discordance in Adjudication Variables:

<table>
<thead>
<tr>
<th></th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Justifiable Difference in Lesion Selection</td>
</tr>
<tr>
<td>2</td>
<td>Incorrect Lesion Selection</td>
</tr>
<tr>
<td>3</td>
<td>Justifiable Perception Difference in Determining New Lesions</td>
</tr>
<tr>
<td>4</td>
<td>Incorrect Perception Difference in Determining New Lesions</td>
</tr>
<tr>
<td>5</td>
<td>Justifiable Perception Difference in Determining Progression on the Basis of Non-target Disease</td>
</tr>
<tr>
<td>6</td>
<td>Incorrect Perception Difference in Determining Progression on the Basis of Non-target Disease</td>
</tr>
<tr>
<td>7</td>
<td>Justifiable Perception Difference in Lesion Measurements</td>
</tr>
<tr>
<td>8</td>
<td>Incorrect Perception Difference in Lesion Measurements</td>
</tr>
<tr>
<td>9</td>
<td>Missing Clinical Data</td>
</tr>
<tr>
<td>10</td>
<td>Quality Issues</td>
</tr>
</tbody>
</table>
• Quality control chart used to determine if a business process is within statistical control.

• Specifics of when to use a “P Chart”:
  – Counting outcomes of an event class (adjudication is the event class)
  – An event has exactly 2 possible outcomes (accept/reject)
  – Data is collected in subgroups which may be of varying sizes (readers who have read different numbers of subjects)
  – Takes into account the distribution of acceptance rates across readers and the number of subjects evaluated by each reader in setting upper and lower limits
Corporate Level P Chart

29 Protocols, 3,944 Adjudications, 24 Readers

“Win/Loss Rates” per Reader

- **X-axis:** 24 Reviewers
- **Y-axis:** Adjudication Acceptance Rates

- **Blue Dots:** Adjudication Acceptance Rates per Reviewer
- **Green Lines:** Warning Limits
- **Red Lines:** Action Limits

- R3 & R10 below lower action limit - investigation & corrective action required
- R7, R21 & R22 above upper action limit – no action required (overachievers)

© 2010 Confidential and proprietary property of CoreLab Partners, Inc, all rights reserved.
• Readers qualified at corporate level and monitored on per protocol basis
• Perhaps all readers being “inside the box” for the protocol is more important than the overall adjudication rate.
Site vs. Central Review
Site vs. Central Discordance

• Outcome from site is different than the Blinded Independent Central Review (BICR)

• Is this really discordance?
  – The process is very different
  – BICR process well controlled (trained readers, Quality metrics)
  – The data may be different at site (prior exams, clinical information)
  – Who is selecting/measuring lesions at sites (reader training)
  – Bias (ICL is Blinded to Tx arm and investigator input)
  – Review P Chart Analysis data for Reader Adjudications from BICR

• Or are these just differences, which are understandable, and possibly useful for monitoring sites with high discordance rates relative to BICR data
SUMMARY

1. Consider making the Imaging Charter a required document to be submitted with protocol for FDA approval for studies that use imaging as an endpoint.

2. Require monitoring of reader metrics so that adjudication rates are better understood.

3. Possibly use site/central discrepancy data to evaluate sites with a high discordance rate relative to Blinded Independent Central Review data.