The Case Against Blinded Review of Microscopic Specimens for Qualification of Novel Biomarkers

Karamjeet Pandher BVSc, PhD, DACVP
Drug Safety Research and Development
Pfizer Inc.
Disclaimer

Opinions expressed are those of the presenter and do not reflect the opinions of Pfizer Inc.
Blinded Review of Microscopic Specimens for Qualification of Biomarkers:

• Has been a topic of discussion for at least 30 years

• Blinded review in toxicology studies is not endorsed by the major veterinary pathology professional organizations

• Best practices:
    – carefully compare treated and controls
    – understand dose response
    – define terminology, severity scores
    – targeted blind read as needed
    – peer-review

• Biomarker studies should be evaluated by deploying nothing less

• Pathologist will be blinded to test biomarkers
Blinded Review of Microscopic Specimens for Qualification of Biomarkers: cons

• When is blinded review adequate?
  – Syndromes where histologic changes and diagnostic criteria are well-characterized e.g., Mankin score, grading for NASH
  – Targeted blind review
  – Quantitative analysis

• Not adequate/detrimental when trying to define a lesion and understand new syndromes: best practices should be followed
  – Biomarker studies will involve both classic and novel toxicants: the first aspect of morphologic analysis might represent a learning phase in which diagnostic criteria are established
  – Current study design for biomarkers stipulates comprehensive histopathologic analysis, not analysis of a specific syndrome
Blinded Review of Microscopic Specimens for Qualification of Biomarkers: cons

• In the absence of clearly defined criteria and lack of concurrent controls there is potential for increased operator bias.
  – Pathologist will be dependent on historical knowledge and experience
  – The suggested two tier approach: one set of slides of learning and one for data recording
  – Objectives achieved by targeted blind read

• Blinding may not be feasible in all studies e.g., toxicant (ANIT) resulting in multiple morphologic manifestations
Assessment of single cell necrosis.
Histopathology and Biomarker Performance

• Receiver operating characteristic (ROC) curve
  – Will be used to categorize biomarker values as true or false positive
  – Histopathology data will be binary (0 or 1, 2, 3, 4)

• In this light let’s analyze data by House et al,
  – Demonstrated a simple bias in a t-test comparison of blind vs. open analysis by a single operator.
  – Bias was determined in necrosis but not degeneration
<table>
<thead>
<tr>
<th>Score for reading</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>degeneration</td>
</tr>
<tr>
<td>0</td>
<td>102</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
</tbody>
</table>
Even though statisticians may be biased to think that blinding is the best practice for any scientific data, **it might not always be the best approach or might not even be possible:** the opinion of subject matter experts should be taken into consideration.
Some might argue that there is no such thing as a “gold” standard. Even pathologic analysis results are not 100% accurate because, like radiology, pathology is an interpretative discipline. For all studies it is important to have operational standards that take into account the condition being studied, the objectives of the study, and the potential effects of any bias. Some common sense is needed as well.

Thank you
A solution to imperfect standard bias is to apply one of several statistical corrections. To apply these corrections, one must make some assumptions about the imperfect reference standard (e.g., that its sensitivity and specificity are known) and/or the relationship between the results of the test being assessed and the results of the reference-standard test (e.g., that the test in question and the reference-standard test make errors independently of one another). There is continuing research of new statistical methods for addressing imperfect standard bias.