Claims Guidance Panel

Nancy Obuchowski, moderator
Michael Boss
Andrew Buckler
Grace Hyun Kim
Paul Kinahan
Nicholas Petrick
Outline

1. Presentation:
   a) Examples of claims
   b) Statistical assumptions
   c) Steps in developing claim

2. Questions/Discussion
Currently, QIBA has two types of claims:

• **Cross-sectional**: measurement at a single time point

• **Longitudinal**: change in measurement over two time points
Patient-Centric

• The claim language is patient-centric, describing the quantitative interpretation of the measurements for the individual patient (or feature of a patient).

• If Profile were to be used in a clinical trial, you’d need information about between-subject variability to plan sample size for trial.
Example of Cross-Sectional Claim

For an ADC measurement of $X \text{ mm}^2/\text{s}$ in solid tumors greater than 1 cm in diameter or twice the slice thickness (whichever is greater), a 95% confidence interval for the true ADC value is

$$X \pm 5 \times 10^{-10}\text{mm}^2/\text{s}.$$
Example of a **Longitudinal Claim**
(two-parts)

- A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 24%.

- The 95% confidence interval for the true change is

\[ Y_1 - Y_2 \pm 1.96 \sqrt{(Y_1 \times 0.087)^2 + (Y_2 \times 0.087)^2} \]

where \( Y_1 \) and \( Y_2 \) are volume measurements at the two time points.
Statistical Assumptions
Cross-Sectional Claim

For an ADC measurement of $X \text{ mm}^2/s$ in solid tumors greater than 1 cm in diameter or twice the slice thickness (whichever is greater), a 95% confidence interval for the true ADC value is

$X \pm 5 \times 10^{-10}\text{mm}^2/s$.

Assumptions:

1. Measurements are normally distributed
2. Bias can be measured
3. Measurement error ("$5 \times 10^{-10}$") includes both bias and imprecision
4. Measurement error is constant
Statistics of Cross-Sectional Claims

Assumptions:

2. Bias can be measured

- If bias cannot be measured, weaker claims can be made
  a. Can claim how repeatable the measurements are
  b. Can claim how well measurements discriminate between disease and non-disease
Statistics of Cross-Sectional Claims

Assumptions:

3. Measurement error includes both bias and imprecision

\[ \text{total error} = \text{bias}^2 + w\text{SD}^2 \]
Statistics of Cross-Sectional Claims

Assumptions:

4. Measurement error is constant

• Bias and wSD are constant for all ADC values.
• If not true, then:
  a) Determine if % bias and wCV are constant
  b) Provide multiple claims
Example of Multiple Claims

• Claim 1: A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 24% and the longest in-plane diameter is initially 50-100mm.

• Claim 2: A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 29% and the longest in-plane diameter is initially 35-49mm.

• Claim 3: A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 39% and the longest in-plane diameter is initially 10-34mm.

• Claim 4: The 95% confidence interval for the change is

\[ Y_1 - Y_2 \pm 1.96 \sqrt{(Y_1 \times wCV_1)^2 + (Y_2 \times wCV_2)^2} \]
Statistics of Longitudinal Claims

The assumptions differ depending on whether different imaging machines/readers are used at the two time points.
Statistics of Longitudinal Claims with same imaging machine/reader

Assumptions:
1. Measurements are normally distributed
2. Measurement precision is constant
3. Linearity

Note that bias does not need to be measured.
Statistics of Longitudinal Claims with same imaging machine/reader

Assumptions:

2. Measurement precision is constant
   • If not true, then:
     a) Determine if wCV is constant
     b) Provide multiple claims
Statistics of Longitudinal Claims with same imaging machine/reader

Assumptions:

3. Linearity (preferably with slope=1)
Statistics of Longitudinal Claims with different imaging machine/reader

Assumptions:
1. Measurements are normally distributed
2. Bias can be measured
3. Measurement precision is constant
4. Linearity holds
Steps in Developing Claim
Steps in Determining Performance Value used in Claim

<table>
<thead>
<tr>
<th>Steps</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Choose appropriate statistical metric</td>
</tr>
<tr>
<td>2</td>
<td>Determine characteristics which degrade performance</td>
</tr>
<tr>
<td>3</td>
<td>Identify plausible set of performance values</td>
</tr>
<tr>
<td>4</td>
<td>Consider clinical requirements</td>
</tr>
<tr>
<td>5</td>
<td>Consider sample size for testing compliance</td>
</tr>
<tr>
<td>6</td>
<td>Choose performance value</td>
</tr>
</tbody>
</table>
1. Cross-sectional or Longitudinal Claim?

2. Is the wSD or wCV constant?
3. Is there bias in the measurements?

- **Scenario A**
  Constant wSD; negligible bias: Use wSD to construct 95% CI

- **Scenario C**
  Constant wCV; negligible bias: Use wCV to construct 95% CI

- **Scenario E**
  Multiple wCVs; negligible bias: Multiple claims using different wCVs to construct 95% CIs

- **Scenario B**
  Constant wSD; bias is common: Use TDI with wSD to construct 95% CI

- **Scenario D**
  Constant wCV; bias is common: Use TDI with wCV to construct 95% CI

- **Scenario G**
  Constant wSD: Use wSD to estimate $R_C^*$ and construct 95% CI

- **Scenario H**
  Constant wCV: Use wCV to estimate $R_C^*$ and construct 95% CI

- **Scenario J**
  Constant wSD; negligible bias: Use wSD to estimate $R_D$ and construct 95% CI

- **Scenario K**
  Constant wCV; negligible bias: Use wCV to estimate $R_D$ and construct 95% CI

- **Scenario M**
  Constant wSD; bias is common: Use TDI with wSD to estimate $R_D$ and construct 95% CI

- **Scenario N**
  Constant wCV; bias is common: Use TDI with wCV to estimate $R_D$ and construct 95% CI

- **Scenario O**
  Multiple wCVs; bias is common: Multiple claims using different TDIs to estimate RDCs and construct 95% CIs

- **Scenario I**
  Multiple wCVs; bias is common: Multiple claims using different TDIs to estimate RDCs and construct 95% CIs
Step 2: Determine characteristics which degrade performance

- Spiculated lesions degrade performance
- Head motion degrades performance

Do you need separate claims for these?
Do you exclude these from the claim?
Step 3: Identify plausible set of performance values

- Some groups have performed literature reviews and conducted meta-analyses to determine the typical bias and precision.
Step 3: Identify plausible set of performance values

- Example: Perc 15 Profile
  4 studies
  summary estimate for wSD was 4.0 HU
  95% CI for wSD was [1.7, 6.6]
Step 3: Identify plausible set of performance values

- Other groups have performed challenge studies to estimate vendors’ bias and precision in both simulated and clinical models.
Step 3: Identify plausible set of performance values

- Example: CT Volumetry Profile
  10 vendors
  bottom one-third: wCV=18%
  middle one-third: wCV=8-9%
  top one-third: wCV=4-5%
Step 4: Consider clinical requirements

• Perc 15 Profile:
  – 95% CI for wSD was $[1.7, 6.6]$.  
  – Experts believed that a wSD of 6-7 HU is acceptable clinically.

• CT Volumetry Profile:
  – vendors demonstrated wCVs of 4-18%.  
  – Experts judged these to be clinically quite large.
Step 5: Consider sample size for testing compliance

- Perc 15 Profile: Suppose claim uses wSD=7
  - Then for actors with wSD=4.0, only need a sample of $\approx 13$ subjects to show that $wSD \leq 7$.

- CT Volumetry Profile: Suppose use $wCV=8\%$.
  - For actors with $wCV=5\%$, need $\approx 17$ subjects
  - For actors with $wCV=6\%$, need $\approx 43$ subjects
  - For actors with $wCV=7\%$, need $\approx 190$ subjects
Step 6: Choose performance value

- Perc 15 Profile: Experts chose a wSD of 6.5 HU \(\Rightarrow\) Repeatability Coefficient of 18 HU.

- CT Volumetry Profile: Experts chose wCV of 8.7\% \(\Rightarrow\) Repeatability Coefficient of 24\%. 
Questions from Audience