November 17, 2015

Claims Guidance*

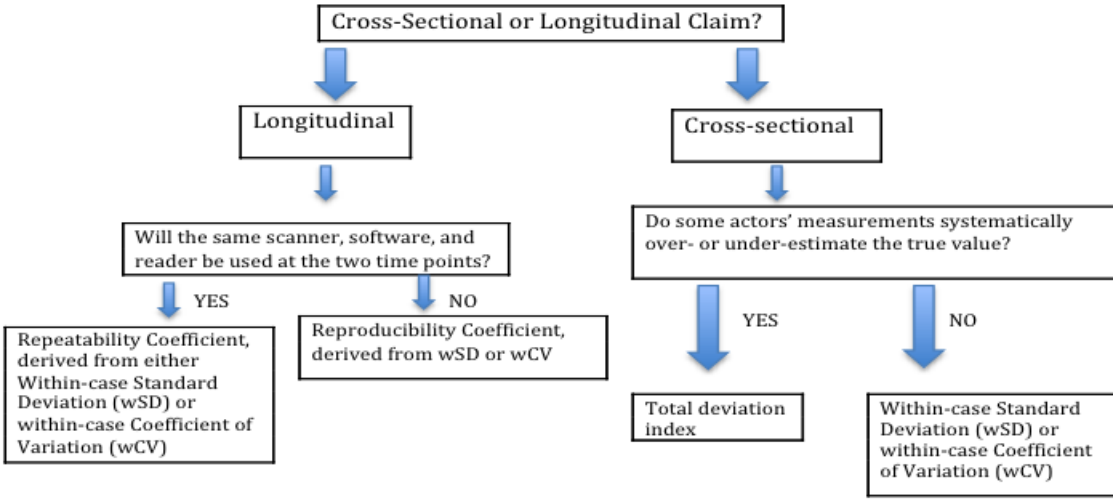
- Claims involve one or more summary statements of the technical performance of the QIB. There are two kinds of claims: cross-sectional and longitudinal. A cross-sectional claim describes the imaging procedure's ability to measure the QIB at one time point, while a longitudinal claim describes the ability to measure the change in the QIB over multiple time points.

- The claim language is patient-centric, describing the quantitative interpretation of the measurements for the individual patient.

- The steps for choosing a precision value for the claim statements are as follows:
    - **Step 1: Choose Metric.** The choice of statistical metrics depends on: (1) whether the imaging biomarker measurements tend to be biased or unbiased, (2) whether the claim is cross-sectional or longitudinal, and (3) whether the precision is constant or varies over the range of measurements. See Figure 1 and examples of metrics below.
    - **Step 2: Determine Characteristics which Degrade Precision.** When technical performance is affected by patient or tumor characteristics, and if these characteristics are prevalent in the general population, then the performance value used in the claim statement is often limited to apply only to the appropriate subpopulations of tumors or patients. For example, spiculated tumors may be more difficult to measure (i.e. result in less precision) than spherical tumors. Center of mass may be measured with less precision in patients with excessive head movement. The claim values need to account for imprecision in measuring the QIB for these characteristics based on their relative prevalence in the population.
    - **Step 3: Identify Plausible Range.** Data from published papers and groundwork projects are used to estimate a range of field performance values. This range might be the 95% confidence interval (CI) of the performance from a meta-analysis of published studies. Alternatively, this range might be based on results from groundwork projects in QIBA or conducted by another outside group.
    - **Step 4: Consider Clinical Requirements.** When available, the clinical needs for the QIB performance are considered. For example, we ask: How small does tumor perfusion change need to be before medication is changed? How precise does the volume of a lung nodule need to be estimated so suspicious nodules are appropriately biopsied and stable nodules are followed? When possible, these clinical needs are considered in determining the performance value for the claim.
    - **Step 5: Consider Sample Size for Conformance Test.** Whereas many of the requirements documented in the Profile are declaratory in nature, a subset of the requirements need to be demonstrated by a given actor which seeks to indicate that they conform. If an actor's imaging device has

precision very close to the required performance value, then very large studies are needed to verify that the actor's imaging device conforms with the requirement. If an actor's imaging device has performance much better than the required performance value, then smaller studies could be adequate.

- **Step 6: Choose Performance Value.** From the plausible range in step 3, and taking into consideration the clinical needs and sample size requirements for testing conformance in steps 4-5, experts from the fields of imaging physics and medicine choose a reasonable performance value for the Profile.



Cross-sectional claims should use the following style:

"For a QIB measurement of Y units, a 95% confidence interval for the true QIB value is Y ± precision value."

- Example 1 (Constant SD): *"For an ADC measurement of X 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 ± 5 $\times 10^{-10} mm^2/s$."* Note that "*5 $\times 10^{-10} mm^2/s$*" is equal to (1.96 × wSD), where wSD is the within-tumor standard deviation (2.55 × 10^{-10} here) and 1.96 is the 95% confidence

factor. It is assumed that the wSD is constant over the range of relevant ADC values.

- Example 2 (Constant wCV): *"For a measured lung tumor volume of Y mm^3, a 95% confidence interval for the true volume is $Y \pm (1.96 \times Y \times 0.14)$."* For some QIB measurements, such as tumor volumes, the precision varies with the magnitude of the measurement. In these cases, precision is often quantified by the wCV (wSD/Y). In this example the wCV=0.14 (or 14%). It is assumed that wCV is constant over the range of relevant tumor volumes.

- Example 3 (Look-up Table for wCV): *"For a measured lung nodule volume of Y mm^3, a 95% confidence interval for the true volume is $Y \pm (1.96 \times Y \times wCV)$."* For some QIB measurements, such as tumor nodules, not only does the precision vary with the magnitude of the measurement, but we cannot assume that the wCV is constant. In these situations a look-up table is provided in the Profile which lists the wCV for various ranges of the measured QIB. The user must use the table to determine which wCV should be used based on the measured Y.

- Following each claim statement, there should be footnotes which describe
  - the statistical metric used in the claim,
  - the statistical assumptions underlying the claim, and
  - realistic examples illustrating use of the claim.
  - For example, one might say, "These claims are based on estimates of the within-tumor coefficient of variation (wCV) for nodules in this size range. In the claim statement the CI is expressed as $Y \pm 1.96 \times Y \times wCV$. The claim is based on the assumption that the wCV is constant for tumors in the specified size range and that there is negligible bias in the measurements (i.e. bias < 5%).

Longitudinal claims should use the following two-part style:

"A measured change in the QIB of Δ or larger indicates that a true change has occurred with 95% confidence"

and

"For a measured change of Δ, a 95% confidence interval for the true change is $\Delta \pm$ precision value."

- Example 1 (Constant RC): *"A measured decrease in Perc15 of 18 HU or more without volume adjustment indicates that a true increase in the extent of emphysema has occurred with 95% confidence. For a measured change of Δ HU in Perc15 without volume adjustment, a 95% confidence interval for the true change is [Δ -18 HU, Δ +18 HU]."* Note that "18" is the Repeatability Coefficient, or ($1.96 \times \sqrt{(2)} \times$ wSD). It is assumed that the wSD is constant over the range of relevant Perc15 values.

- Example 2 (Constant wCV): *"A measured change in the tumor's volume of* $\Delta\%$ *indicates that a true change has occurred with 95% confidence if* $\Delta\%$ *is larger than 38%"* and *"If* Y_1 *and* Y_2 *are tumor volume measurements at the two time points, a 95% confidence interval for the true change is* $(Y_2 - Y_1) \pm 1.96 \times \sqrt{(Y_1 \times 0.14)^2 + (Y_2 \times 0.14)^2}$. For some QIB measurements, such as tumor volumes, the precision varies with the magnitude of the measurement. In these cases, precision is often quantified by the wCV (wSD/Y). In this example, the wCV=0.14 (or 14%). Then the RC is $(2.77 \times \text{wCV} \times 100) = 38\%$. It is assumed that wCV is constant over the range of relevant tumor volumes.

- Example 3 (Look-up Table for wCV): *"A measured change in the QIB measurements of* $\Delta\%$ *indicates that a true change has occurred with 95% confidence if* $\Delta\%$ *is larger than (2.77* $\times$ *wCV* $\times$ *100)"* and *"If* Y_1 *and* Y_2 *are the QIB measurements at the two time points, a 95% confidence interval for the true change is* $(Y_2 - Y_1) \pm 1.96 \times \sqrt{(Y_1 \times wCV)^2 + (Y_2 \times wCV)^2}$." For some QIB measurements, such as tumor nodules, not only does the precision vary with the magnitude of the measurement, but we cannot assume that the wCV is constant. In these situations a look-up table is provided in the Profile which lists the wCV for various ranges of the measured QIB. The user must use the table to determine which wCVs should be used based on the measured Y_1 and Y_2.

- Following each claim statement, there should be footnotes which describe
    - the statistical metric used in the claim,
    - the statistical assumptions underlying the claim,
    - the imaging methods used at the two time points, and
    - realistic examples illustrating use of the claim.
    - For example, one might say, "These claims are based on estimates of the within-nodule coefficient of variation (wCV) for nodules in this size range. For estimating the critical % change, the % Repeatability Coefficient (%RC) is used: $2.77 \times \text{wCV} \times 100$. The claim is based on the assumptions that the same imaging methods will be used at the two time points, the wCV is constant for nodules in the specified size range, and that the measurements follow the linearity property with slope equal to one (i.e. slope differs from unity by < 5%).

* Most of this work can be found in "Statistical Issues in Testing Conformance with the Quantitative Imaging Biomarker Alliance (QIBA) Profile Claims", Obuchowski et al. Under review *Academic Radiology*.