March 9, 2021

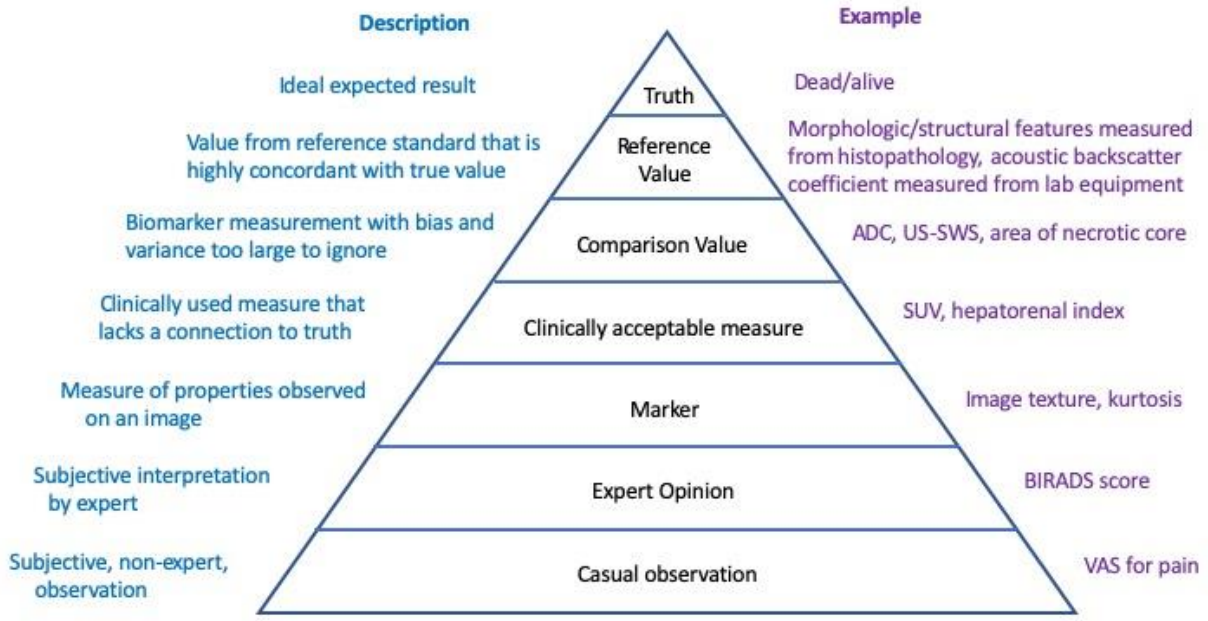
Recommendations of the Ground Truth Task Force

The Ground Truth Task Force was formed in December 2020 by QIBA leadership to assess the role of "ground truth" in QIBA's Profiles. The Task Force met three times over January and February.

Following Sullivan et al [2015], the Task Force chose not to use the term "ground truth" but rather discussed "true value", "reference value", and related measurements. In their first meeting the Task Force established the need to define a hierarchy of terms that included "true value", "reference value", "biomarker measurements", etc. to delineate the relationships between these terms. The hierarchy was defined and revised during the second and third meetings. Once the hierarchy was established, the Task Force used it to provide recommendations to QIBA regarding the role of true and reference values in their Profiles.

Staying consistent with descriptions of "truth", "reference standard", and "reference value" provided in Sullivan et al [2015], the hierarchy depicted in the following figure was established. At the top is "true value", followed by measurements of decreasing levels of certainty. A brief description of each level is provided on the left, with examples on the right. The Task Force made the following observations regarding this hierarchy:

1. The True Value, although not always easily observed, is assumed to exist.
2. Reference Value and Comparison Value are each defined relative to a true value. The pyramid levels below these may not have a relationship to a true value.
3. Reference Values have negligible bias and imprecision, such that they can be considered highly concordant with the True Value [Sullivan et al, 2015].
4. Biomarkers fall under Comparison Values where there is a clear connection to truth. Practically, though, a true or reference value may not be available for all biomarkers. When available, measurement bias can be estimated relative to the true or reference value.
5. When comparing a biomarker to another Comparison Value, agreement can be assessed, but the bias of the biomarker cannot be estimated.
6. Standardization of imaging methods relative to the true value is possible with biomarkers but not with Clinically Acceptable Measurements and Markers. The latter should still be standardized against a state-of-the-art Clinically Acceptable Measurement or Marker, when available.
7. Expert Opinion and Casual Observation are typically not quantitative in nature and thus are not the focus of QIBA.



Task Force Recommendations: The ability to characterize and quantify the bias of a measurement relative to a true or reference value is important to quantitative imaging. This is possible with biomarkers, but not 'Clinically Acceptable' measures, nor imaging 'Markers'. QIBA has focused on biomarkers in the past and has established a respected and trusted methodology for writing Profiles for biomarkers. In contrast, 'Markers' lack a connection to truth; thus, performance metrics currently used by QIBA, such as bias and linearity, as well as quantification of true change over time, are not applicable to 'Markers'. The Task Force does not feel that imaging 'Markers' should categorically be excluded from QIBA; however, given limited resources and the fundamental differences in performance metrics, the Task Force recommends that QIBA prioritize Profiles involving biomarkers over Profiles involving markers.

True or reference values are available for many QIBA biomarkers, allowing estimation of measurement bias. An example is tumor volume measurements on CT, where phantoms can be constructed such that the true volume is known. Another example is ADC where a true value is available from an ice-water phantom. For other biomarkers a true or reference value is not available or is too expensive, and thus measurement bias cannot be estimated. For example, ultrasound shear wave speed and magnetic resonance imaging elastography are biomarkers where there is a clear connection to the concept of a shear modulus for any material, but no reference material or method exists for shear modulus in the range of soft tissues (~1kPa). In the past QIBA has included biomarkers where no reference value is available, focusing on longitudinal claims in the Profiles. The Task Force does not see a need to change this approach.

Multiparametric quantitative imaging (mp-QI) is of great interest to the clinical community, as well as to QIBA. The Task Force recommends that QIBA prioritize biomarkers in their mp-QI

Profiles, over markers. Furthermore, since calibration is an important metric for describing the performance of clinical prediction models, QIBA should prioritize models whose outcome has a relationship with the true value so that the calibration of the model can be reported.

Reference:

Sullivan et al, 'Metrology Standards for Quantitative Imaging Biomarkers', Radiology 2015

Task Force Members:

Andrew Buckler, MS
Jana Delfino, PhD
Alexander Guimaraes, MD, PhD
Tim Hall, PhD
P. David Mozley, MD
Nancy Obuchowski, PhD (Chair)
Nicholas Petrick, PhD
Gudrun Zahlmann, PhD