

In This Issue:

IN MY OPINION

The Importance of Achieving Claim Confirmed and Clinically Confirmed Profile Status

By RICHARD L. WAHL, MD, FACR

ANALYSIS: TOOLS & TECHNIQUES

Stage 4: Claim Confirmed – Why Do We Need This?

By NANCY OBUCHOWSKI, PhD

QIBA IN THE LITERATURE

QIBA MISSION

Improve the value and practicality of quantitative imaging biomarkers by reducing variability across devices, sites, patients, and time.

QIBA CONNECTIONS

[QIBA Wiki](#)

QIBA CONTACT

[Contact Us](#)

Edward F. Jackson, PhD

QIBA Chair

In My Opinion

The Importance of Achieving Claim Confirmed and Clinically Confirmed Profile Status

By Richard L. Wahl, M.D.

Radiology is gradually moving from a qualitative and subjective field to a quantitative and objective one. While qualitative image assessments and synthesis of clinical knowledge will remain key components of interpretation of images, the inherently quantitative nature of many of our imaging methods has led to greater emphasis on quantitation. However, accurate quantitation requires attention to detail in the imaging procedures and many steps along the way can go awry. Colloquially, if the quality of the acquired imaging data is not consistent, one can be in a “garbage in/garbage out” situation regarding quantitative data. Indeed, I have often discussed with my trainees that “bad quantitation is worse than no quantitation.”

One of the most quantitative and understood imaging methodologies is PET/CT most commonly performed using the radiotracer FDG. FDG PET/CT is performed in several million patients each year globally, most commonly in patients with cancer or suspected cancer, but often interpreted qualitatively.

FDG PET has been one of the early areas of focus in the QIBA Profile development process. While it would seem quite simple, the evolution of QIBA and of the FDG Profile maturation has been gradual. But the Profile development has steadily and logically led to a rather mature Profile which is ready for broad utilization pending additional validation at many sites, using a process as outlined by Nancy Obuchowski, PhD, in this newsletter (1).

The QIBA Profile development process is shown below:

Stage 1: Public Comment	<i>The Biomarker Committee experts have drafted the Profile and believe it is practical and expect it to achieve the claimed performance.</i> (Status: Done for FDG)
Stage 2: Consensus	<i>The wider community has read the Profile and believe it to be practical and expect it to achieve the claimed performance.</i> (Status: Done for FDG)
Stage 3: Technically Confirmed	<i>Several sites have performed the Profile and found it to be practical and expect it to achieve the claimed performance.</i> (Status: Done for FDG)
Stage 4: Claim Confirmed	<i>Some sites have performed the Profile and found it achieved the claimed performance.</i> (Status: Substantially Complete for FDG)
Stage 5: Clinically Confirmed	<i>Many sites have performed the Profile and demonstrated the claimed performance is widely achievable.</i>

Achieving a Practical Profile

The hundreds of volunteers who make up the critical mass of QIBA along with RSNA liaisons, have made considerable progress in developing clinically relevant and workable Profiles, including for FDG PET/CT. Since FDG PET/CT is already in clinical use and many oncologists request that Standardized Uptake Values (SUV) be reported in routine reports, it is worth considering how a clinically confirmed QIBA Profile may impact clinical trials and clinical practice once *Claim Confirmed stage* (Stage 4) is fully achieved.

Some practical lessons have evolved from the QIBA Profile maturation status from Stage 1 to early Stage 4. First, in early stages of Profile development, there

is a tendency to attempt to achieve a “perfect” Profile. However, to achieve perfection (or nearly so) there was great complexity, and some aspects of the Profile were viewed as unworkable or not fully understood. Thus, there was a quick recognition that “perfect was the enemy of the good,” and that a Profile which could not be performed except at the most advanced imaging science centers with an army of physicists/case, would likely not be a workable approach.

This is why the word “practical” appears in the description of Stages 1-3 as an important element. If many clinical sites cannot perform the Profile, a determination has to be made as to whether the Profile is too complex, or if there are tools that need to be put in place by manufacturers so the key process can be undertaken. This is why the review by a broad range of experts in the field and pilot field testing in Stage 3 are/were so important. A Profile that is too complex to perform adds little immediate value to quantitative imaging.

What is Stage 4: *Claim Confirmed*?

This stage is simply an independent validation of the assumptions underlying our claim statements as discussed by Dr. Obuchowski. Is the performance we are claiming actually met in the field? While we already know of several sites using the Profile who have achieved the Profile claim, we do need a breadth of performance sites using varying instrumentation for image acquisition and software to independently verify that their site can achieve the performance claim using the Profile.

Indeed, early experience in Stages 1-3 led us away from an “absolute SUV measurement claim,” since we are not yet certain if scanners in different sites and created by different manufacturers with varying software and reconstructions will give us “identical” SUV’s if measured on the same individual with cancer who has been injected with FDG, or even on the same phantom. However, we have focused our claim on the ability of a given scanner at a specific site to be able to achieve nearly the “same” SUV max measurement in a test/re-test setting in which a patient with cancer, not receiving treatment, would be expected to have a similar SUV max if the test were repeated in a few days after the first scan.

How will achieving *Claim Confirmed* and *Clinically Confirmed* status empower physicians?

QIBA Profiles are designed to inform clinical trials and clinical practice. In clinical trials, it is possible that more resources can be identified by the trial sponsor to validate scanner performance and possibly to provide central data analysis and review. Such carefully designed studies may meet or exceed the performance claim but may have a data intensity monitoring process more robust than achievable in clinical practice.

Determining within a trial if a change in SUV over time during therapy is significant is important if imaging response adapted therapies are to be undertaken. For example, an early PET assessment, after just one cycle of therapy, may be used as an integral biomarker to determine if a patient should stay on therapy, have therapy intensified, or therapy de-intensified early in the therapy.

Using an early version of a QIBA Profile, we have shown that the early changes in FDG uptake during Primary Systemic Therapy of breast cancer with chemotherapy or monoclonal antibodies is highly predictive of eventual pathological outcomes ^(2,3). Knowing what degree of change of SUV is significant is critical in driving such studies, so changes in treatment, if informed by PET, are based on real, as opposed to measurement variability-related, changes in SUV max.

Other examples of use of change in SUV measurements with treatment can include studies to examine early anti-tumor effects of therapies of varying doses and combinations. Knowing what degree of change is “real” vs. chance measurement error, is very important in driving sample sizes and in modern clinical trial design. It is totally feasible that the QIBA FDG Profile will be able to drive dose de-intensification for excellent responders, for example.

While quantitation may not be needed at the end of therapy to determine if a response is complete, for these early responses, the quantitative metrics are of critical importance. A systematic measurement approach like that outlined in PERCIST 1.0 may further inform reporting of SUV changes, as our Profiles evolve to assess metrics like SUV-lean peak and measures of tumor volume. We are

pleased that our efforts to validate the QIBA Profile for FDG PET/CT in our clinic have shown performance in the expected range for the existing Profile, supporting the performance of this Profile (4).

For clinical practice, it is clear that change in patient management usually will require a definite change in cancer metabolic status. With the FDG PET Profile eventually completing Stages 4 and 5, with all assessments and analyses done on site (and not at a central lab), we expect that a practicing physician ordering or interpreting an FDG PET/CT from a clinic following the QIBA Profile, can have greater confidence that an increase or decrease in SUV max from a baseline metric, exceeding the threshold defined in the Profile, is very likely to indicate a real biological change in tumor status.

In the presence of therapy, this will often be a treatment-related effect. The change required to be medically significant may differ from the change required to be statistically significant (for example a decline SUV of a breast cancer of 50% after 6 cycles of therapy may not be a very good response, but this same decline after 10 days of therapy may predict a meaningful improvement in patient outcome if chemotherapy was applied).

Simply being able to reliably advise our referring physicians, based on solid evidence if a change in SUV max is “real” or not will greatly inform our practice and help better bring quantitative imaging into the mainstream of PET/CT interpretation. Indeed, it is completely possible that an eventual “quality metric” for FDG PET might be the adherence to QIBA Profiles and providing quantitative results. Obviously, this should be evidence-based, but the evidence to date suggests greater agreement among readers of PET using quantitation than of those using qualitative metrics.

As Dr. Obuchowski emphasizes, “the results from the study should be generalizable to a broad spectrum of sites, patients, and imaging methods.” With such knowledge, we will have a *Claim Confirmed* and ultimately a *Clinically Confirmed* QIBA Profile. This will allow clinical trials and clinicians to have confidence that they have achieved robust quantitation and can use quantitative FDG PET/CT in their clinical trials and clinical practice. In this way, the quantitative in vivo phenotyping available from FDG PET/CT can drive

modern clinical trials and better inform clinical practice, resulting in true precision medicine driven substantially by quantitative FDG PET/CT imaging.

References:

- [1] Stage 4: Claim Confirmed – Why do we need this? Nancy A. Obuchowski, PhD
Cleveland Clinic Foundation (this newsletter).
- [2] Connolly, RM, et al. TBCRC 008: Early change in 18F-FDG uptake on PET predicts response to preoperative systemic therapy in human epidermal growth factor receptor 2–negative primary operable breast cancer. *Journal of Nuclear Medicine*, 2015;56:1;31-37.
- [3] Connolly, RM, et al., TBCRC026: Phase II clinical trial assessing the correlation of standardized uptake value on FDG PET/CT with pathological complete response to pertuzumab and trastuzumab in primary operable HER2-positive breast cancer. *Journal of Clinical Oncology* 2019;36;511)
- [4] Fraum, TJ, et al, Measurement Repeatability of 18F-FDG-PET/CT versus 18F-FDG-PET/MRI in Solid Tumors of the Pelvis. (*Journal of Nuclear Medicine*, 2019; In Press.
- [5] O, JH, et al, Response to early treatment evaluated with 18F-FDG PET and PERCIST 1.0 predicts survival in patients with Ewing sarcoma family of tumors treated with a monoclonal antibody to the insulin-like growth factor 1 receptor. (*Journal of Nuclear Medicine* 2016;57:5;735-740.

PMR



Figure: Representative FDG PET changes in Sarcoma treated with anti IGF1R antibody. Visual changes are striking, but quantitative SUV change data have substantial predictive value for survival. O, JH, et al, Journal of Nuclear Medicine 2016;57:5;735-740.



Richard L. Wahl, MD

Richard L. Wahl, MD, is the Elizabeth E. Mallinckrodt Professor and Chair of Radiology at Washington University School of Medicine, St. Louis, and Director of the Mallinckrodt Institute of Radiology. His research interests include the use of quantitative imaging, particularly FDG-PET, to quantify and predict the response of cancers to therapy as well as to guide cancer theranostics. He is a member of the QIBA Nuclear Medicine Coordinating Committee and has been actively engaged in developing the QIBA FDG PET/CT Profile.

Analysis: Tools and Techniques

Stage 4: Claim Confirmed – Why Do We Need This?

By Nancy A. Obuchowski, PhD

QIBA has made excellent progress in developing clinically relevant and workable Profiles for many imaging biomarker applications. At the time of this writing, QIBA has 5 Profiles at the *Consensus* stage (Stage 2) and 2 at the *Technically Confirmed* stage (Stage 3).

While Profiles are available and usable at any stage, a *Claim Confirmed* Profile (Stage 4) will have the biggest impact on clinical use and adoption into clinical trials. In this article I discuss the objectives of the *Claim Confirmed* stage and how we can achieve it.

What is Stage 4: Claim Confirmed?

The *Claim Confirmed* stage is simply an independent validation of the assumptions underlying our claim statements. For example, all of our claims rely on estimates of repeatability in one way or another. In a Stage 4 study, we would independently estimate repeatability (e.g. test-retest variance) and compare the estimate with the value used in the claim statement.

Hopefully, our estimate matches up with the value used in the claim (i.e. statistically speaking, the upper 95% confidence bound should be \leq the value used in the claim; if not, we need to adjust the claim).

What kind of study is needed?

Of all the previous stages, the *Claim Confirmed* stage might be the most challenging. It requires new data to be collected and processed under the "*Technically Confirmed*" Profile. The key requirements for the study are applicability, generalizability and statistical power.

By applicability, I mean that the study should include an assessment of all key assumptions, and the imaging methods need to strictly follow the Profile. Table 1 summarizes the statistical assumptions underlying different types of claims. Simpler claims require a simpler study, while more complex claims may require multiple studies.

The results from the study should be generalizable to a broad spectrum of sites, patients, and imaging methods. This usually means accrual of patients through a multi-site design from both academic and private institutions. The patients need to represent the spectrum of patient characteristics described in the Profile. The sites should be chosen to represent different vendors. If the biomarker estimation requires a human reader, the performance of different readers needs to be assessed. A core lab could be used, particularly if the biomarker will be used most often in clinical trials; however, there is a loss of generalizability with a core lab. Some investigators have proposed processing the scans at both a core lab, as well as at the clinical sites where the images were obtained.

Validation studies do not need to be large. Ideally, a few cases should be accrued from each of 5-6 sites, rather than a large number of cases from 1-2 sites. For example, to validate a claim based on an assumed wCV of 10%, 30 test-retest subjects are needed (for a study with 80% power, 5% one-sided test, and when the true wCV is $\leq 8\%$). This could mean just 5-6 subjects per site. For some biomarker applications, subjects may have multiple lesions; inclusion of multiple lesions from the same subject reduces the overall sample size somewhat, though the inter-lesion correlation must be accounted for. It's important not to rely on too many measurements from the same patient; a rule of thumb is <5 lesions per patient.

Table 1: Requirements for Stage 4: Claim Confirmed Studies

Type of Claim	Underlying Statistical Assumptions	Type of Study for Validation
Performance claim about repeatability	Estimate of wSD or wCV ¹ [1]	Test-retest study on human subjects [2]
Cross-sectional claim	Estimate of wSD or wCV and Estimate of bias [1]	Test-retest study on human subjects and phantom study ²
Longitudinal claim (same imaging methods at two time points)	Estimate of wSD or wCV (If claim includes 95% CI for magnitude of the change, then also need to assess linearity and estimate slope [1].)	Test-retest study on human subjects (phantom study ²)
Longitudinal claim (different imaging methods at two time points)	Estimate of wSD or wCV, Linearity and estimate of slope, and Estimate of bias	Test-retest study on human subjects and phantom study ^{2,3}

¹within-subject standard deviation (wSD), within-subject coefficient of variation (wCV)

²A phantom study (or clinical study with reference standard, if available) can be designed to assess linearity and estimate the slope of a regression of the biomarker measurements against the ground truth values [2, 3].

³The reproducibility of the biomarker measurements can be estimated either directly (from a test-retest study with patients imaged on two different scanners) or indirectly through estimation of the bias and repeatability.

Why do we need to do this?

Most of the statistical data used to inform our Profile claims comes from meta-analyses of studies in the literature or more seldom from groundwork studies. These data are not collected strictly under the specifications of the Profile because it is still being developed. Furthermore, a lot of expert opinion is required to determine which data/estimates should be used for formulating the claims. Thus, an independent, objective validation is essential.

Once we have a *Claim Confirmed* Profile, it allows other sites to have confidence in the claim statements without having to perform the validation on their own. I feel this is the impetus for achieving this stage, i.e. the claims are validated and ready to be applied wherever the Profile is adopted.

References:

- [1] Kessler LG, et al. The emerging science of quantitative imaging biomarkers: terminology and definitions for scientific studies and for regulatory submissions. *Stat Meth Med Res.* 2015; 24:9-26.
- [2] Obuchowski NA, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Meth Med Res.* 2015; 24: 68-106.
- [3] Raunig D, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Meth Med Res.* 2015; 24: 27-67.



Nancy Obuchowski, PhD, is Vice-Chairman of Quantitative Health Sciences at the Cleveland Clinic and Professor of Medicine at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University. She is a Fellow of the American Statistical Association. Her research interests include study design and statistical analysis methods for imaging screening and diagnostic tests and imaging biomarkers. She is a member of the QIBA Steering Committee

Nancy A. Obuchowski, PhD



QIBA Activities

QIBA Biomarker Committees are open to all interested persons. Meeting summaries and other documents are available on the QIBA website RSNA.ORG/QIBA and wiki <http://qibawiki.rsna.org/>.

QIBA Resources:

[QIBA Webpage](#)
[QIBA Wiki](#)

[QIBA Biomarker Committees](#)
[QIBA Organization Chart](#)

Please contact QIBA@rsna.org for more information. We welcome your participation.

[QIBA and QI/Imaging Biomarkers in the Literature](#)

This list of references showcases articles that mention QIBA, quantitative imaging, or quantitative imaging biomarkers. In most cases, these are articles published by QIBA members or relate to a research project undertaken by QIBA members that may have received special recognition. New submissions are welcome and may be directed to QIBA@rsna.org.