QIBA Newsletter

April 2013 • Volume 5, Number 2

QIBA MISSION

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

QIBA CONNECTIONS

QIBA Wiki

QIBA CONTACT

Contact Us

Daniel C. Sullivan, MD
RSNA Science Advisor

IN MY OPINION

Assessing the Measurement Variability of Lung Lesions in Patient Data Sets (1B Study)—A Researcher's Perspective: the Problem of Intra-reader Variability

By P. DAVID MOZLEY, MD

A landmark study by QIBA's CT Volumetry Technical Committee[1] last year produced some startling results: intra-rater variability can exceed 150% for measurements of tumor diameters and 800% for the corresponding volumes. This extreme variability threatens the integrity of the fundamental method for assessing responses to treatment in patients with solid tumors.

A review of the 1B study design shows how this variability happened and demonstrates how to prevent it from happening again. The specific aim of the study was to determine the minimum level of change that can be reliably measured with current technology. The dataset[2] was acquired under no-biological-change conditions by scanning patients twice after a short time interval. Target lesions were pre-selected by the 1B team. In the first phase of the study, each image of the pair was presented for measurement as a single time point to experienced radiologists in the imaging core lab industry. After several weeks, the second image was presented as if it was a separate case in order to promote independent intra-reader assessments. In the second phase of the study, the analysts were allowed to see their work on the companion time point. To mitigate the risk that the raters would realize the pairs represented no change, the cases were mixed with new cases that actually did reflect changes.

Findings showed that the highest intra-reader variability occurred in cases with substantial inter-reader variability as well. Disagreements between experts who, as individuals, demarcated the edges inconsistently, implies that the readers were uncertain about where the tumor boundaries were actually located. In both clinical practice and clinical trials, seasoned veterans will not select target lesions they do not think they can measure with confidence. With that in mind, this study suggests that image analysts should be empowered to select their own target lesions. Frequent concerns about the practice leading to target lesion selection bias can be offset by the reality that variability can explode when readers are forced to measure pre-selected masses. Findings also suggest that the drive to force measurements of every tumor mass in an effort to quantify the whole body tumor burden could be precarious in some contexts. The will to measure every tumor needs to be tempered by the risk of ruining measurements that are sound with guesses about others that are susceptible to extreme variability.

Misconceptions Plague Measurability

Unexpected variability prompted the 1B team to retrospectively commission an assessment of each lesion for its measurability. An independent expert from the imaging core lab industry concluded that a substantial minority of these tumors were not measurable. Some experts from academia disagreed, pointing out that if these lesions were not classified as measurable, patients would not be eligible to participate in clinical trials that require subjects to have at least one measurable lesion.

In my opinion, misconceptions about what is measureable still plague the field and constitute one of the major causes of discordance between site and central assessments. Selecting masses with visible boundaries on only a few central slices risks making a measurement in areas that do not reflect the aggressiveness of the neoplasm. The most proliferation can occur on the edges that are the hardest to see; that is, the zone of most uncertainty can sometimes correspond to the zone of most interest. Reliably measuring inactive tissue is vain. Until better segmentation algorithms are developed, the safest approach is to treat most of these masses as non-target lesions. While it is true that such a sad salute to reality would diminish the objectivity of the measurements, the independent assessment of the objective radiological evidence would not be compromised, and evidence that can be classified as "purely objective" might be viewed with less skepticism.

During the first phase of the 1B study, the image analysts were not informed of a companion time point or allowed to see their prior work on it. Not reviewing a prior study would be classified as gross negligence in clinical practice. In clinical trials, best practices usually obligate reviewers to see their prior time point assessments, as well as the impact of their new boundaries on the quantitative change measures before locking the current time point. This "instantaneous reality check" constitutes a quality control procedure that tends to constrain intra-rater variability. Comparing the intra-reader variability in the first and second phases of the 1B study produced evidence supporting the soundness of this workflow. For years, many well-meaning scientists and regulatory officers have advocated presenting follow-up scans in random order. The 1B team showed this will confound the assessments too often to be a viable workflow in many trials of new treatments for cancer.

Implementing the lessons learned from the 1B study should reduce intra-reader variability to random noise. In conducting clinical trials, truly coping with random noise requires only money. Any problems can be solved by constructing a trial with adequate statistical power, because random noise tends to affect both arms of a properly designed trial equally, introducing no bias. As such, variability is simply another business expense. More noise means more subjects per trial, and as a consequence, more money spent on accrual, patient care, etc. But, after all, it's only money.

For these reasons, it's my opinion that the QIBA 1B team demonstrated that intra-rater variability doesn't have to be a serious problem in clinical research.

Editor's note: See below to access data from QIBA's CT Volumetry Technical Committee 1B Study. A peer-reviewed manuscript is currently being prepared.

REFERENCES:
[1]

http://qibawiki.rsna.org/images/d/d6/QIBA_Exp1B_CoffeeBreak_summary_Table_26Jan11_RECIST_Measureable_or_NOT.pdf

[2] https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections

P. David Mozley, MD, helped design and execute clinical trials that used quantitative imaging as an endpoint for 11 years in the pharmaceutical industry before returning to academia this past year. He represented the Pharmaceutical Imaging Group on the QIBA Steering Committee and co-chaired the QIBA CT Modality and Technical Committees from 2008-2012.

Each issue of *QIBA Newsletter* features a link to a dynamic search in PubMed, the National Library of Medicine's interface to its MEDLINE database. Link to articles on: "Assessing the Measurement Variability of Lung Lesions in Patient Data Sets (1B Study)—A Researcher's Perspective: the Problem of Intra-reader Variability."

ANALYSIS: TOOLS & TECHNIQUES

Development of a DWI Phantom for Quantitative ADC Measurement

MICHAEL A. BOSS, PH.D.

There has been an increased interest in recent years in acquiring apparent diffusion coefficient (ADC) maps for disease diagnosis and treatment monitoring. In many tumor models, an increased cellularity leads to a decrease in tissue water mobility, providing a contrast mechanism between pathologic and healthy tissues. During treatment, there is a tendency for this cellularity to decrease as pathologic tissue dies; with increasing necrosis, water mobility increases. Thus, many tumors can be detected by way of a decreased ADC relative to surrounding healthy tissue and the efficacy of treatment determined by monitoring changes in ADC.

The QIBA Perfusion/Diffusion/Flow MRI (PDF-MRI) Technical Committee recently began developing a Profile for diffusion-weighted MR imaging (DW-MRI), focusing on obtaining an ADC map and designing a new ADC phantom for quality control purposes. Previous ADC phantoms have made great strides in characterizing scanner performance, especially regarding repeatability and reproducibility, best exemplified by an ice water phantom developed at the University of Michigan in conjunction with the National Cancer Institute.

The ice water phantom has provided valuable data about scanner performance in terms of bias and variance.[1] Using an ice bath, temperature can be maintained at 0 degrees Celsius, mitigating changes in diffusion due to thermal effects. The diffusion coefficient of water at 0 degrees Celsius is well known, and serves as a ground truth for determination of bias. In a multicenter trial, the ice water phantom revealed that most MR imaging scanners measure ADC with a repeatability of approximately 2% at the isocenter of the magnet. However, significant bias was noted when measuring ADC off-isocenter.[2]

These lessons have proven useful in the next generation of ADC phantom design. The diffusion coefficient of water molecules can be tuned using mixtures of water and polyvinylpyrrolidone (PVP); varying concentration in a multi-compartment phantom allows for a spread of ADC values approximating those seen in vivo, as demonstrated in a prototype PVP phantom (Figure 1). Temperature control will be achieved using ice water, or temperature could be measured using the chemical shift of protons on a lanthanide chelate such as Tm-DOTMA or via MR-compatible fiber optic thermometers. This is necessary given the range of temperatures found in scanner rooms and the associated changes in the true diffusion rate. Spatial dependence of ADC within the scanner will be ascertained by means of strategically placed cells off of isocenter within the phantom (Fig. 2). Ultimately, the ADC phantom will enable better quality control and properly address measurement issues to allow for quantitative imaging.
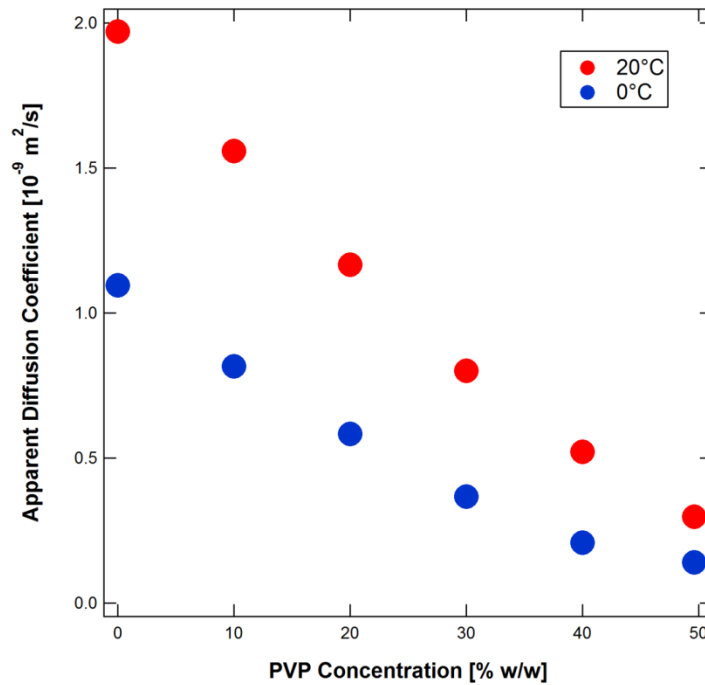
Figure 1. *ADC as a function of PVP concentration. At room temperature, the PVP solution covers a biologically-relevant range of ADC values. The temperature-sensitivity of ADC is readily apparent in going from 20 °C to 0 °C. Data courtesy of T.L. Chenevert.*

Figure 2. *Conceptual design of the multi-compartment ADC phantom. Cells containing different concentrations of PVP near magnet isocenter allow for a range of ADC values to be measured. Strategically placed cells allow for determination of measurement bias off of isocenter. The phantom can be filled with ice water to control temperature, or can accommodate MR-compatible temperature sensors.*

REFERENCES:

[1] Chenevert, TL, Galbàn, CJ, et al., J. Mag. Reson. Imag; 2011 34:983-987. DOI: 10.1002/jmri.22363

[2] Malyarenko, D, Galbàn, CJ, et al., J. Mag. Reson. Imag.,2012; DOI: 10.1002/jmri.23825

Michael A. Boss, PhD, is an affiliate of the National Institute of Standards and Technology through the University of Colorado at Boulder. He is editor of the Diffusion-Weighted MR Imaging Profile of the QIBA Perfusion/Diffusion/Flow Technical Committee. His primary interests are in developing quantitative MR imaging standards, MR relaxation mechanisms and novel contrast agents.

FOCUS ON

QIBA/NIBIB Groundwork Projects

Over the course of the two-and-a-half year contract between RSNA/QIBA and the National Institute of Biomedical Imaging and Bioengineering (NIBIB), 26 QIBA projects across different imaging modalities were conducted, including the establishment of digital reference objects (DROs) and assessing intra/inter-reader variability using various measurement techniques. Ongoing QIBA projects continue and U.S. Food and Drug Administration (FDA) biomarker qualification efforts are under consideration.

QIBA AND QI/ IMAGING BIOMARKERS IN THE LITERATURE

This list of references showcases articles that mention QIBA, quantitative imaging, or quantitative imaging biomarkers.

QIBA in the Literature

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