Title

Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs

Permalink

https://escholarship.org/uc/item/8wc7j4jt

Authors

Keenan, Kathryn E
Biller, Joshua R
Delfino, Jana G
et al.

Peer reviewed

Recommendations Towards Standards for Quantitative MRI (qMRI) and Outstanding Needs

Kathryn E Keenan, PhD[1], Joshua R Biller, PhD[1], Jana G Delfino, PhD[2], Michael A Boss, PhD[1,3], Mark D Does, PhD[4], Jeffrey L Evelhoch, PhD[5], Mark A Griswold, PhD[6], Jeffrey L Gunter, PhD[7], R Scott Hinks, PhD[8], Stuart W Hoffman, PhD[9], Geena Kim, PhD[10], Riccardo Lattanzi, PhD[11], Xiaojuan Li, PhD[12], Luca Marinelli, PhD[13], Gregory J Metzger, PhD[14], Pratik Mukherjee, MD, PhD[15], Robert J Nordstrom, PhD[16], Adele P Peskin, PhD[17], Elena Perez[18], Stephen E Russek, PhD[1], Berkman Sahiner, PhD[2], Natalie Serkova, PhD[19], Amita Shukla-Dave, PhD[20], Michael Steckner, PhD[21], Karl F Stupic, PhD[1], Lisa J Wilmes, PhD[22], Holden H Wu, PhD[23], Huiming Zhang, PhD[16], Edward F Jackson, PhD[24], Daniel C Sullivan, MD[25]

[1]Physical Measurement Laboratory, National Institute of Standards and Technology, Boulder, CO

[2]Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD

[3]Department of Physics, University of Colorado, Boulder, CO

[4]Vanderbilt University Institute of Imaging Science, Vanderbilt University, Nashville, TN

[5]Merck Research Laboratories, West Point, PA

[6]Department of Radiology, Case Western Reserve University

[7]Departments of Radiology and Information Technology, Mayo Clinic, Rochester, MN

[8]GE Healthcare, Milwaukee, WI

[9]Rehabilitation Research and Development Service, Department of Veterans Affairs, Washington, DC

[10]College of Computer & Information Sciences, Regis University, Denver, CO

[11]Department of Radiology, New York University School of Medicine, New York, NY

[12]Program of Advanced Musculoskeletal Imaging (PAMI), Cleveland Clinic, Cleveland, OH

[13]GE Global Research, Niskayuna, NY

[14]Department of Radiology, University of Minnesota, Minneapolis, MN

[15]Department of Radiology, University of California San Francisco, San Francisco, CA

[16]National Cancer Institute, Bethesda, MD

[17]Information Technology Laboratory, National Institute of Standards and Technology, Boulder, CO

[18]Arterys, San Francisco, CO

Corresponding author: Kathryn Keenan, NIST, 325 Broadway, MC 686-08, Boulder, CO 80305, kathryn.keenan@nist.gov.

[19]Department of Radiology, Anschutz Medical Center, Aurora CO

[20]Departments of Medical Physics and Radiology, Memorial Sloan Kettering Cancer Center, New York, NY

[21]Canon Medical Systems USA, Mayfield Village, OH

[22]Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA

[23]Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA

[24]Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, Madison, WI

[25]Department of Radiology, Duke University Medical Center, Durham, NC

RECOMMENDATIONS

Magnetic resonance imaging (MRI) has made it possible to image soft tissues within the human body, allowing non-invasive determination of information. Quantitative MRI (qMRI) can be defined as the extraction of a characteristic from an MR image that has a magnitude that can be expressed as a number with units (e.g., distance) or relative to a reference material (e.g., proton density as % of water signal). This definition includes measurement of length and volume; relaxation properties (T_1, T_2, T_2*); flow; phase and more. Quantitative MRI has the potential to make a great clinical impact on diagnostics by enabling earlier detection of disease, complementing or replacing biopsy, providing clear numeric differentiation of disease states, and increasing the quality of information available to artificial intelligence algorithms. However, the incredible amount of variability in clinically-used image acquisition and post-processing techniques hinders current efforts to extract reliable, consistent and accurate quantitative information from routine MRI exams. Reference materials and reference objects (phantoms), along with associated image acquisition protocols and software for image and data analysis can be developed to address some of these obstacles. Each of the components: reference object, acquisition protocol and software, should be developed with the challenges of the particular anatomy and method being considered. The reference objects and test methods need to be transferred into the clinic, such that reliable and reproducible qMRI becomes routine. Then, the value of qMRI can make a definitive impact.

The National Institute of Standards and Technology (NIST) hosted workshops in 2014 and 2017 with participants from several organizations working towards standards in quantitative MRI (Table 1). Here we summarize the recommendations of the workshop.

While there is a great deal of interest within the clinical community regarding qMRI, all workshop attendees agreed that clinical adoption of quantitative imaging metrics will only be successful if methods are reliable, repeatable, accurate, and, perhaps most importantly, have an impact on the decision-making process for patient care. To ensure repeatable and accurate methods, reference objects and analysis software should be developed concurrently with the development of the quantitative imaging method. A recent example includes an anthropomorphic, prostate, body-imaging reference object(1) developed by the University of Minnesota in collaboration with High Precision Devices, Inc (Boulder, CO). The UMN/HPD body phantom (Figure 1) uses materials for T_1 and T_2 relaxation times and ADC values that were previously used in the International Society of Magnetic Resonance in Medicine (ISMRM)/NIST MR system phantom and the Quantitative Imaging Biomarkers Alliance (QIBA®)/National Institutes of Health (NIH)/NIST isotropic diffusion phantom, to support the broader translation of a prostate cancer prediction model based solely on qMRI data(2). Additionally, for ease-of-use by clinical sites, development of a given reference object should include associated analysis software packages or a digital package akin to an application programming interface (API) that would include locations of the relevant regions of interest and reference property values (e.g., T_1 relaxation times). This "full package" for method development can ease the transition from research to clinic for qMRI applications.

Reference objects can serve multiple important roles in the development and deployment of a qMRI method. The most common role for a qMRI phantom is as a reference sample that can be imaged to measure and compare the performance of a combined imaging and analysis method across instruments and over time. Some standardized reference objects exist(3–12), as described later in this manuscript, but there is a need for more types of phantoms to evaluate the performance of different types of sequences and MRI contrasts, such as physiologically relevant values of T_1, T_2 and ADC in a single voxel, multi-modality (e.g., MR-PET), magnetization transfer (MT), and others listed in Table 2. Reference objects may also serve important roles in the design of quantitative imaging methods from acquisition strategy through analysis pipeline. First, ex vivo tissues, animal models, tissue extracts, or chemical solutions may be used as phantom representations of in vivo human tissue. Experimental studies of such phantoms have helped establish current qMRI tissue models, and as we aim to probe ever more specific tissue characteristics with qMRI, corresponding experimental studies of suitable tissue phantoms will be needed. Finally, in silico phantoms can be used to evaluate the propagation-of-errors in qMRI methods, considering error from image noise, as well as from both the variance and bias of parameter constraints or assumptions, such as deviations from the assumption of homogeneous and static radiofrequency (RF) fields. For example, the MR extended Cardiac-Torso (MRXCAT) phantom for realistic simulation of cardic MR(13) and the Radiological Society of North America (RSNA) QIBA® digital reference objects (DRO) for dynamic contrast-enhanced (DCE) MRI(14). These computational phantom studies are necessary to define potential performance of a qMRI method and can serve to inform on both model and pulse sequence design. While this manuscript focuses on reference objects for use in comparing acquisition data from various imaging platforms to identify and mitigate sources of bias and various to the degree possible and to define confidence intervals on measures obtained from the acquired data, the workshop attendees appreciated that the data analysis packages may also

contribute significantly to measurement bias and variance. Application-specific DROs allow the assessment of such sources of error for a given analysis package as well as levels of variance across analysis packages and across time. Finally, at the workshop, several advanced reference objects were discussed, and needs identified by participants. We summarize those suggestions in Table 2.

BACKGROUND

In light of the many positive contributions qualitative MRI already makes, existing workflows and decades of experience with image interpretation, a legitimate first question regarding *quantitative* MRI is "Why do we need qMRI?". One answer lies in the difference between clinical and sub-clinical presentation of symptoms – or, the absolute minimum deviation of a physiological or anatomic target from normal. A concrete example would be assessing the presence of a tumor, visible by eye (clinical presentation) vs. detecting the presence of that same tumor earlier as a small collection of cells with abnormal shape or physiology (sub-clinical presentation). The treatment options and patient outcome are typically better for the latter scenario than the former. However, to excel at identifying sub-clinical presentation of disease requires numbers (quantitative analysis) in addition to images.

As a case study, consider the effect of quantitative analysis on the determination of pediatric blood lead levels over the past 60 years. Prior to 1960, the prevailing sentiment was typified by toxicologists like Dr. Robert Kehoe who had "never seen a case of lead poisoning with lead level < 80 μg/dL"(15). New analytical methods, especially those based on atomic absorption spectroscopy, were developed in the 1960s and allowed scientists and clinicians to begin addressing the relationship between blood lead level and toxic manifestations quantitatively. Documented method (quality) control, interlaboratory comparisons and improved instrument performance led to agreement between experts on measured lead values and reduction of the measurement error. Without quantitative laboratory tests, Dr. Kehoe relied on clinical manifestations like colic, encephalopathy or death, which occur when lead blood concentrations exceed 50 μg/dL. Implementation of quantitative analytical techniques have now shown specific sub-clinical manifestation of lead concentration in the blood with deleterious impact on hemoglobin synthesis (40 μg/dL), vitamin D metabolism (30 μg/dL), nerve conduction velocity (20 μg/dL), and IQ level, hearing and growth at values as low as 10 μg/dL(16). Recent reviews have even suggested there is sufficient evidence to lower the Centers for Disease Control (CDC) limit for lead in blood to 2 μg/dL(17).

Until 1960, leading clinicians and toxicologists were certain they had an excellent understanding of the lead toxicology and epidemiological impact. However, only after the application of advanced analytical instrumentation in a standardized way did the community see how little they understood. We may ask the same question about the way MRI is currently practiced. What new patterns and sub-clinical presentations of disease states will reveal themselves, once we have removed (through quality control and standardization) variation in image information and quality due to scanner, site and operator influence? Can we use earlier knowledge of disease state, acquired with qMRI, to treat earlier and therefore more efficiently or effectively? Could we avoid unnecessary treatment or interventions with

improved quantitative metrics to better risk stratify patients? Many experts believe the answer to these questions is unequivocally "Yes".

Lead levels are one example of a biomarker, which is a defined characteristic that is measured using an assay, as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention(18,19). The term "biomarker" is often assumed to imply a measurand of a laboratory test (e.g., blood cholesterol levels), but it can also refer to a clinical measurand like blood pressure or the output of a clinical imaging scan. Imaging methods that provide information about the nature and amount of matter or activities present can be considered biomarkers and are conceptually similar to laboratory assay targets. Image-based biomarkers can be considered as a type of *in vivo* assay. An assay is a procedure for measuring the presence, amount, or functional activity of a target entity (the analyte, measurand or target of the assay). The analyte can be a drug, a biochemical substance, or a structure or process in an organism or organic sample. A biomarker is a characteristic (of normal or disease), and an assay is the test that measures some parameter that reflects the "amount", extent, or degree of biomarker that is present. Common terminology, procedures, and methods have become established in medicine to describe, evaluate and validate laboratory assays through the work of the Clinical and Laboratory Standards Institute (CLSI). Many of the same concepts and approaches can and should be applied to radiological diagnostic assays, and this has begun to occur in an organized way over the past decade through the efforts of many groups (Table 1).

The use of imaging-based biomarkers for clinical applications, drug development and safety evaluation has increased considerably in recent years and has stimulated a large number of efforts to develop methods for accuracy and consistency across medical imaging platforms, especially in the setting of multi-center trials. The National Institute of Standards and Technology hosted workshops in 2006(20), 2014 and 2017 to drive discussion and collaboration amongst the various stakeholders and groups working on standardization for quantitative medical imaging through tools such as reference objects and protocols. Standards range from common understanding or common practices to common requirements, and standards can be established through consensus groups such as the RSNA QIBA® or National Electrical Manufacturers Association (NEMA) (Table 1). A documentary standard is a classification, guide, specification, or test method developed and established according to principles of consensus, such as those of NEMA. Here, we review some of the achievements of the standardization efforts of the past decade and discuss some needed steps to overcome the obstacles inhibiting clinical diagnostics based on qMRI.

DEPLOYED NIST REFERENCE OBJECTS

NIST created three reference objects that have been commercialized and are being actively used by the MRI community: the ISMRM/NIST system phantom(7,21), the QIBA®/NIH/NIST isotropic diffusion phantom(3,22,23), and the University of California San Francisco (UCSF)/NIST breast phantom(4,5). The NIST reference objects have been used to identify measurement errors and then improve the stability of acquisitions in technique/pulse sequence development and protocol development for multi-site studies. These reference objects were necessary to aid standardization, but they are insufficient given the wide range

of qMRI applications – that is, there is still a need for additional reference objects. For example, the breast phantom was created in part because the existing phantoms were not physically compatible with a breast coil.

In general, the reference object (or phantom) is an outer shell in an idealized geometry (e.g., a sphere for a head), which contains reference solutions in specified configuration. The outer container is dependent on both the anatomy of interest and the RF hardware in use (e.g., a 15.0 cm diameter sphere is not a good match for a breast coil). Previous work demonstrated that using phantoms with correct geometry for the anatomy of interest, which is also appropriately sized for the RF hardware, is important for assessing quantitative imaging performance(5,24). The reference solutions within the phantom are dictated by the quantitative parameters of interest. For example, the UCSF/NIST breast phantom contains fat-tissue mimics, while the other reference objects do not.

The ISMRM/NIST system phantom was designed to address many of the needs of qMRI including volumetric distortions, slice profile assessment, high contrast spatial resolution, proton density and T_1 and T_2 relaxation times. The system phantom design includes important features from the Alzheimer's Disease Neuroimaging (ADNI) MagPhan phantom (The Phantom Laboratory, Salem, New York, USA) used to assess geometric distortion in three dimensions (Figure 2) (25,26) and the American College of Radiology (ACR) phantom used to assess slice profile and resolution(27,28). The system phantom reference solutions have various T_1 and T_2 properties that have been fully characterized using NIST traceable methods(29). This phantom has been, for example, used to assess day-to-day variation of MR fingerprinting (MRF) measurements of T_1 and T_2 relaxation times(30). In this study, the ISMRM/NIST system phantom revealed variation in T_2 measurement, which could be a result of temperature variation (such measurements when using $MnCl_2$ solutions are known to be temperature sensitive). However, this finding also led to improvements in the T_2 measurement technique, addressing measurement error due to transmit B_1 variations and reducing measured variation.

Based on the work of Pierpaoli *et al.* to develop aqueous solutions of polyvinylpyrrolidone (PVP) (22,23) the National Institutes of Health (NIH) and Chenevert *et al.* ice-water phantom(31,32), Boss *et al.*(3) developed an isotropic diffusion phantom (Figure 3). The QIBA®/NIH/NIST isotropic diffusion phantom, filled with ice-water, was used by the TRACK-TBI trial to qualify 19 clinical sites as enrolling centers. Across the 11 initial sites, for the 0 to 40 % weight-by-weight (w/w) PVP concentrations, the coefficient of variation increased from 2.2 % to 4.5 %. The 50 % w/w PVP sample had a considerably higher coefficient of variation, 11.9 %, likely due to the chosen protocol that used *b*-values of 0, 500 and 900 s/mm^2(33).

Finally, the UCSF/NIST breast phantom, which has T_1 and ADC values representative of fibroglandular and fat tissue, is used by the I-SPY 2 clinical trial(5). The phantom is designed to test both sides of breast coils, which requires a unique geometry. An initial study with the phantom revealed geometric distortions in the right-left orientation of the axial plane with echo planar diffusion-weighted images (Figure 4) across systems (fields and platforms) dependent on which side of the coil the object was located(4). The distortions

remained in the right/left orientation regardless of phase-encoding direction, did not change with b-value, and thus are most likely due to B_0 inhomogeneity(34). The B_0 inhomogeneity can be due to the magnet itself; the RF or gradient coil hardware; and objects within the scanner (e.g., the phantom or foreign objects in the body). In addition, the distortions were present in patient data as well and were not identified prior to using the phantom with the geometric distortion array. The geometric distortion, which varies both day-to-day and scanner-to-scanner, limits the precision with which tumor sizes can be measured and as a result limits the ability to identify changes in tumor size due to intervention (e.g., radiation or chemotherapy).

The NIST reference objects have a demonstrated benefit for qMRI protocol development across manufacturer systems and an ability to find technical errors in system software and hardware. Reference objects should be used not just as a tool to qualify sites and qualitative data for clinical trials, but additionally to qualify systems to provide qMRI data in order to demonstrate the potential added value and for ongoing quality control programs. Other reference objects that are used by the MR community include the ACR phantom(27,28); structural brain imaging phantom developed by ADNI(26); a dynamic contrast-enhanced MRI perfusion phantom developed by RSNA QIBA®(35); other isotropic diffusion phantoms(31,32); static tissue phantoms for phase-contrast flow applications(36–38); cardiac tissue phantoms(39); $MnCl_2$ phantoms for iron measurement applications(40,41); and proton-density fat fraction phantom(42–44). Finally, the three reference objects reviewed here and those mentioned in the previous sentence are not able to address all the needs of the MRI community. In the sections that follow, we highlight some outstanding needs as identified at the workshop.

STAKEHOLDER PERSPECTIVES FROM THE 2017 WORKSHOP

During the 2017 workshop, several stakeholders presented their perspectives centered around the question "How is qMRI being used and what are the associated needs or challenges?" Summaries of these perspectives are provided below.

Present Status of Standards (Perspective from Organizations Developing Technical Standards)

Documentary standards for MRI are established by several organizations, some of which are based in the United States and some internationally. Documentary standards are beneficial because a proven, robust, well-documented and relatively simple method is identified that is known to produce a reliable output. Documentary standards for image performance (e.g., SNR, uniformity), ghosting, resolution and concept of quality assurance testing are developed and maintained by the NEMA Medical Imaging and Technology Alliance (NEMA/MITA) as found in the NEMA MS series 1–12(45–47) and the International Electrotechnical Commission (IEC) (62464–1)(48). Safety (e.g., whole body, local specific absorption rate (SAR); acoustic noise) standards for MRI are developed and maintained by both the IEC as IEC 60601–2–33(49) and some standards in the NEMA MS series. Currently, adoption of all MRI test standards is voluntary in the United States, and these standards may be used by MRI system manufacturers when obtaining marketing

authorization, assessing safety, or performing basic system quality assurance tests. The Digital Imaging and Communications in Medicine (DICOM) standard image file format is also maintained by NEMA/MITA. To receive reimbursements from Medicare, MRI systems, at least at outpatient facilities, must annually pass a Medicare-approved certification process (such as, but not limited to, the ACR MRI Accreditation Program)(6,50). However, the tests in these Medicare-approved accreditation processes are only partially quantitative and do not include evaluation of T_1 or T_2 relaxation times or ADC measurements.

Documentary standards may be a useful mechanism for enabling a unified approach to specific qMRI methodologies. However, prior to creating new documentary standards or tests, we caution the community to ask: "what new information does an additional test or standard deliver?". If the goal is an imaging protocol that works across systems, a documentary standard may not be the best way to achieve that goal. Without agreement on one method to make a measurement, guidelines or recommendations, such as those promulgated by RSNA QIBA® (Table 1), may be the best path toward meeting the community's needs for standardized acquisition protocols.

Members of the community can lead the adoption of de facto standards by championing their own cause. A necessary step is to collect and publish test-retest data that leads to identifying repeatable protocols and the creation of a standardized protocol across manufacturers' systems. For example, the magnetic resonance elastography (MRE) community conducted several test-retest studies(51–53) and leveraged those to create a QIBA® profile for performing MRE in the liver(54). With non-trivial effort, the quantitative MRI community itself can lead the development and adoption of standardized practices. However, consensus agreement must still be reached across the community to support a particular standard.

In addition to the work of QIBA®, other efforts toward consensus methods come from specific communities. For example, the cardiac MR community regularly publishes consensus methods(55–57), which are available at https://scmr.org/page/guidelines. ISMRM organized or endorsed workshops led to consensus statements on diffusion outside the brain(58), fat-water standardization(59,60), speech MRI(61), and a proposed standard for raw data format(62). The ISMRM perfusion study group along with the European consortium for Arterial Spin Labeled (ASL) in dementia wrote a consensus statement on recommended implementation of ASL perfusion MRI(63), which led to inclusion in the ACR practice parameters documentation(64). As discussed earlier in this section, a standard should be adopted with consensus and because it addresses an unmet need or leads to the sharing of new information.

Seeing the Same: the Need for Platform Independence in Imaging (Perspective from the Veterans Health Administration)

The principal mission of the Veterans Health Administration is the health of the military Veteran in the United States. One example relates to brain disorders. While improvements in outer tactical vests (body armor) and helmets have reduced fatal injuries, many Servicemembers return with a traumatic brain injury (TBI), post-traumatic stress disorder (PTSD), suicidal thoughts or behaviors, and/or related co-morbidities(65). These co-

morbidities or co-occurring conditions are defined as mental health disorders by the National Research Action Plan (NRAP)(66). One of the main objectives of the NRAP is to develop methods that could detect TBI and PTSD in Servicemembers and Veterans and enable the monitoring of progression of the condition and its response to therapy.

MRI could provide safe, noninvasive imaging techniques to meet many of the goals of the NRAP. For example, diffusion tensor imaging (DTI) could be used to visualize white matter tracts, which are known to be susceptible to progressive aspects of TBI(67). However, differences across manufacturers, models, and software versions as well as possible "drift over time" have hindered use of the technology, even for clinical trials(68). Differences across manufacturers, models, post-processing techniques and time were recognized at the workshop as a key hinderance to clinical adoption. Differences exist for simple geometric measurements, as documented by the ADNI trials(26), and for model-based measurements such as T1 relaxation time measurement(69). Reference objects were used to identify the geometric distortions in ADNI trials(26) and assess T1 measurement variation(69). Reference objects are critical to ensure quantitative MRI data is comparable between systems (both hardware and software variations) for neurodegenerative disorders and other applications. Developing DTI into an accurate, reliable diagnostic that can be used to assess the health of white matter over time could lead to a noninvasive tool that not only can track neurodegenerative disease, but also the efficacy of therapies designed to halt progressive degeneration or to repair the damaged brain for Veterans, Servicemembers, and civilians.

Quantitative Imaging Biomarkers in the Era of Precision Medicine (Perspective from Clinical Research)

In the era of precision (personalized) medicine, quantitative imaging biomarkers have become a pre-requisite. Quantitative imaging develops and optimizes anatomical, functional and molecular imaging protocols, data analyses, display methods and reporting structures. Quantitative imaging biomarkers can address unmet medical needs and help in providing personalized medicine.

In MRI, there are several relevant quantitative imaging biomarkers derived from specific MRI techniques, and few have been incorporated in the clinical workflow. One such technique is diffusion-weighted MRI (DW-MRI), which is sensitive to random motion of water molecules. The quantitative imaging biomarker derived from mono-exponential modeling of DW-MRI data is the apparent diffusion coefficient (ADC). In the clinic, ADC has been used to stage tumors(70–73), assess treatment response(74,75), and predict tumor aggressiveness(76,77). However, at present, confidence currently doesn't exist to compare data across scanners and populations, preventing the use of quantitative ADC measures in clinical workflow. Qualification and validation would reduce measurement variability across sites and lead to comparable data across scanners, sites, and populations so that ADC and other biomarkers could be reliably assessed, and data aggregated from multiple clinical trials. Once that step is taken, it is conceivable that ADC could be used to replace biopsy in some cases or better inform the 'watch and wait' path.

Therefore, there is an urgent need to validate and qualify quantitative imaging biomarkers, such as ADC, through a rigorous process, which involves a partnership between manufacturers and academic research centers, such as the work by RSNA QIBA®(78).

Quantitative Assessment of MRI-Guided Interventions (Perspective from Academic Research)

Magnetic resonance imaging (MRI)-guided interventions, also known as interventional MRI (iMRI), takes advantage of the unique strengths of MRI to plan, guide, monitor, and assess diagnostic and therapeutic interventions. In light of the increasing clinical applications of iMRI and continual development of new technologies for iMRI, quantitative assessment is critical to characterize performance, understand limitations, facilitate and benchmark new developments, and ensure procedural safety, efficiency, and success.

To enable quantitative assessment of iMRI, it is informative to classify iMRI procedures as delivery (e.g., of a device, energy, or agent) vs. extraction (e.g., of tissue or a device), or device-based (e.g., targeted needle biopsy) vs. energy-based (e.g., focal ablation). The technical parameters to quantify include: position and geometry of the device/tissue, effects of the device (e.g., image quality, SNR, contrast-to-noise ratio (CNR), geometric distortion, tissue properties, safety), position and amount of the energy/agent (e.g., temperature, concentration), and effects of the energy/agent (e.g., image quality, SNR and CNR, tissue properties, safety). In addition, the application aspects to quantify include: procedural workflow (e.g., number of steps, time per step, failed/repeated steps, repeatability of each step, learning curve and operator dependency), technical success (e.g., accuracy of targeted device placement, accuracy of location and dose of energy/agent delivery), and clinical success (e.g., patient outcomes).

Each of the components for quantitative assessment should be standardized by the community, including scientists, clinicians, and industry experts for specific iMRI applications. As an example, the American Association of Physicists in Medicine (AAPM) Task Group on MRI-Guided Robotics-Assisted Interventions is developing standardized 3D-printed reference objects, experimental configurations, MRI protocols, performance metrics, and analysis software to quantify the impact of MRI-guided robots on image SNR and geometric distortion, as well as the accuracy and precision of MRI-guided robotic-assisted targeted device placement. Previous efforts have developed an MRI-compatible motion platform to emulate respiratory motion(79) for application in dynamic MRI(80) and real-time MRI-guided procedures(81) (Figure 5).

The impact of robotic systems and interventional devices on MRI SNR can be determined by measuring SNR (according to ACR and NEMA guidelines) with and without the device and then calculating the percentage difference. While it is obviously desirable to design systems and devices that do not degrade SNR, acceptable levels of SNR percentage difference should be considered in the context of specific clinical applications. Similarly, the impact of devices on geometric distortion can be assessed by imaging a phantom with known structural features (e.g., a grid plate) with and without the device and then calculating the observed mis-alignment of features on MRI (e.g., displacement in mm and rotation in degrees). The amount of geometric distortion depends on the choice of MRI sequence (e.g.,

balanced steady-state free precession, gradient echo, or spin echo) and imaging parameters (e.g., readout bandwidth, echo time (TE), repetition time (TR)), as each sequence has a different degree of sensitivity to perturbations caused by devices such as susceptibility gradients and off-resonance effects. The acceptable level of geometric distortion depends on the dimensions of the tissue of interest and the definition of technical/clinical success for the procedure. When possible, the MRI sequence with most appropriate CNR for the anatomy of interest should be used, and the imaging parameters should be prescribed to reduce geometric distortion to acceptable levels. In cases where this is not readily achievable, a second MRI sequence with acceptable levels of geometric distortion can be used to provide complementary information for guidance. Image artifacts due to systems and devices can also be assessed and managed using similar methods. In addition, the technical accuracy and precision of a particular interventional workflow and system, such as in targeted needle placement, can be characterized by performing repeated experiments in a standardized phantom with known target locations and then analyzing the distribution of an error metric (e.g., 3D Euclidean distance to measure needle-to-target error in mm). For energy delivery applications, the temperature measurement accuracy of MRI (in degrees Celsius) could be assessed with respect to a known reference (e.g., stable body temperature in a subject without heating or fiber-optic temperature probes). The acceptable accuracy and precision of the temperature measurement would need to be evaluated in the context of required therapeutic dose for treatment in diseased tissues and safety bounds for avoiding damage to normal tissues.

In summary, quantitative assessment is crucial for the research and application of iMRI. Additional efforts are needed for community building and standardization in iMRI, and there will be much to gain by leveraging and joining related endeavors in diagnostic MRI and MRI safety.

Survey of Current Reference Object Use in Multi-Site Studies (Perspective from Pharmaceutical Research)

The biopharmaceutical industry uses qMRI to guide the determination of drug efficacy and selection. At the level of multi-site trials, imaging clinical research organizations (iCROs) are often engaged to conduct the clinical trials. The iCROs conduct the imaging component of clinical trials in partnership with pharmaceutical companies and contribute to the study design, the training and qualification of sites, and harmonization of protocols, and provide a central image quality control and analysis. For this workshop, Jeffrey Evelhoch informally surveyed five iCROs (BioClinica, ICON, IXICO, PAREXEL and VirtualScopics) on their use of reference objects for MRI clinical trials. In general, when the clinical trial uses qMRI as an endpoint, MRI reference objects are used for site qualification, site monitoring(40,41), QA/QC(6) or as a tissue reference (e.g., fat fraction(42–44)). However, while all of the polled iCROs were aware of the NIST reference objects, none of them used the NIST reference objects for their qMRI trials. There are at least two reasons for that: first, the iCRO customers have not requested the use of NIST reference objects; second, some trials employed qMRI to assess metrics that were not included in the NIST phantoms (e.g., iron). There may be a need for NIST or other national metrology institutes to be involved in the creation of reference objects to contribute expertise, such as the contributions of NIST and

Physikalisch-Technische Bundesanstalt (PTB) for the T_1 Mapping and ECV Standardization (T1MES) cardiac phantom(39).

The Need for Reference Objects in Multi-Site Studies for "Killer Applications" (Perspective from Medical Imaging Research)

Utilization of imaging for neurological disease is evolving rapidly. While structural (qualitative) imaging remains the mainstay of MRI referrals for diseases of the brain, researchers and clinicians are exploring the value qMRI brings to diagnosis, prognosis, treatment optimization and monitoring for patients suffering from neurological disease. Standard of care for TBI today is CT imaging to rule out brain bleeds, which often require immediate surgical intervention. However, CT imaging is normal in the vast majority of mild TBI cases, even for subjects with significant symptoms and who eventually have a complex and prolonged recovery. TRACK-TBI has demonstrated that 3T MRI is more sensitive than CT in discovering soft tissue pathology in the brain following trauma, but even with this sensitive modality, qualitative structural MRI is still of limited use in the clinical management of mild TBI patients(82,83). Several studies have explored the value of qMRI techniques such as diffusion imaging, functional MRI, perfusion arterial spin labeling, susceptibility mapping, MR spectroscopy, and other approaches; however, a complete picture has not yet emerged.

GE Healthcare and the National Football League (NFL) have partnered to study the value of advanced qMRI in a clinical research study in mild TBI(84). To overcome the recognized limitations of existing MR techniques, a state-of-the-art, multi-contrast MRI protocol was developed and distributed to the seven participating sites using uniform scanner hardware (3T MR750 with Nova Medical 32 channel RF brain coil) and software. The same imaging protocol was deployed and monitored at all sites, data was processed by a core lab, and analyzed by a team of GE scientists. Despite these stringent controls, subtle differences were detected across most sites, which could be attributed primarily to differences in the patient populations. One site was recognized as a statistical outlier for quantitative diffusion MRI data analysis. A root cause analysis is in progress to determine the source of these significant but small, quantitative differences that were limited to the high b-value shell and that do not affect routine, clinical imaging. A multi-shell (up to b=800, 1200, 2800 s/mm^2, 147 q-space points distributed on three shells) diffusion imaging protocol was run at all sites(85). Comparison of the data across sites on a single, traveling human phantom showed a coefficient of variation for mean fractional anisotropy (averaged over all whole white matter) of 1.6 % and 2.0 %, and mean orthogonal diffusional kurtosis of 1.9 % and 3.3 % respectively, excluding and including the statistical outlier site. At the time of the data collection, there was no appropriate phantom. Existence of an appropriate phantom for the specific quantitative metric under study (for example, restricted anisotropic diffusion) would improve harmonization of the data across sites as well as a testing paradigm to recognize some of these statistical anomalies.

Even with these technical considerations, qMRI biomarkers were identified and related to clinical measures of severity and progression of the disease. Resting state functional MRI results demonstrated remarkable correlation with clinical presentation and

symptomatology(86,87), perhaps creating an opportunity to replace some of the subjective symptom inventories with objective imaging-based measures of the severity of the disease. Perfusion arterial spin labeling demonstrated decoupling of the clinical and physiological recovery of the brain(88). Symptoms disappeared sooner than physiological blood flow changes and could point to the necessity to re-evaluate return-to-activity criteria in light of the persisting abnormal blood-flow physiology in certain individuals. Finally, diffusion imaging techniques beyond current routinely used diffusion tensor models unlocked additional sensitivity to subtle axonal pathology following trauma. Not only could traumatic axonal injury can be quantified and monitored, but also axonal swelling potentially could be determined via diffusion kurtosis imaging, which characterizes tissue microstructure(89). An outlier analysis compared each individual subject to a population of normal controls and identified regions of the brain where these axonal swelling effects were preferentially noticed(85).

Deep Learning and Medical Imaging (Perspective from Industrial Research)

Deep learning techniques are used in a wide variety of fields for discovery of features in very large data sets. In medical imaging, these techniques help to identify and measure structures within images to follow the progress of disease. In order to use the output of deep learning models to make medical decisions, it is essential to be able to quantify their accuracy. Arterys (San Francisco, CA) presented an update on their deep learning models for cardiac medical images, one of which was approved by FDA for medical use. Arterys' first commercial product is a deep learning model that was trained for segmentation of the left and right ventricles in cardiac MR images. The key to this success lies in careful work to obtain the right data to train, validate, and assess accuracy of the model. Large data sets of annotated images were screened and collected to train their model and address missing data or inconsistencies in the training annotations. Models were validated during construction using a collection of separate images annotated by radiologists. Determining the accuracy required further testing to a reference data set. The model was evaluated on two sets of data: one with hundreds of studies, where the goal was to show $< 10\%$ volume error and another with a few dozen studies, where the goal was to show accuracy within the expected range of expert annotators. By accessing industry-accepted consensus data, Arterys was able to obtain quantitative comparison of accuracy and precision compared to expert annotators. These reference data sets are not always available, yet there is a real need for a consensus ground truth in all applications. The lack of reference data limits the use of deep learning models in medicine, as it becomes difficult to validate algorithms and compare them to physician's output.

Pre-clinical MRI (Perspective from Pre-Clinical Research)

Pre-clinical (animal) scanners, historically, have been subject to fewer regulations and standards than their clinical counterparts. However, animal MRI is considered a quantitative technique since the NIH and other research funding agencies expect to see quantitative biomarkers as the end result of a preclinical study to be readily translated into clinical trials. Most animal studies, similarly to clinical trials, are based on a longitudinal study design with repetitive MRI acquisitions on the same animal. Ensuring that the data obtained are reliable and accurate requires good image quality and quality controls. To achieve reliable results

from pre-clinical scanners and imaging protocols over time, quality assurance programs, industry/manufacturer standards and qMRI reference objects are desperately needed. Even the most sophisticated, high-field pre-clinical MRI scanners are installed and validated with simplified manufacturer phantoms, which are often based on conical 50-cc tubes filled with fluids for contrast (no relaxation time measurements) and small, interlocking, plastic, toy "bricks" for volumetric measurements and anatomical localization (Figure 6). There are only a few published reports on home-built pre-clinical MRI reference objects(90,91), which are assembled without defined standard operating procedures. None of those prototypes have been broadly adopted outside the individual laboratories involved in the design. Nevertheless, these individual efforts show the existing demand for and tendency towards designing and implementing anatomical and tissue specific multi-modal, multi-parametric, preclinical qMRI phantoms(91,92). One suggestion from the workshop was to create an open-source design of a reference object for pre-clinical imaging, such that a site could make a reliable reference object particular to their magnetic field strength and bore and coil sizes. The availability of such phantoms would facilitate the promise of translation to human imaging studies.

DISCUSSION

Several of the perspectives presented here demonstrate that qMRI applications can produce highly valuable measures for applications in TBI, cancer diagnostics and treatment planning, clinical trials and preclinical research. However, technical obstacles can prevent accurate measurements, longitudinal comparisons, and multi-site comparisons for single patients and populations. For example, qMRI measurements are only as good as the model implementation and overly simplified models could result in misleading qMRI results (e.g., assignment of a single $T1$ and $T2$ value to macroscopic voxels that contain a mix of multiple $T1$ and $T2$ values). Clinical and technical validation of qMRI techniques is needed, and the workshop attendees proposed and discussed appropriate best practices and next steps.

The increased emphasis on evidence-based and precision medicine requires physicians to integrate data from clinical examinations, laboratory tests, and imaging studies when deciding on patient care, and to assess and alter the plan as necessary. Such integration of data from multiple sources is becoming increasingly automated, and this requires that input data be interoperable, machine-readable and, ideally, quantitative. Decision-support tools and artificial intelligence algorithms need to perform their tasks using reproducible data. Quantitative MRI procedures must move systematically toward a standardized, reproducible, high-throughput format for clinical diagnostic implementation. Clinical validation builds upon the technical performance of the qMRI method and requires that the technical performance of the test meets the needed performance specifications; poor reproducibility may hinder clinical validation. Similarly, there cannot be a high level of confidence in conclusions drawn from clinical trials that depend on qMRI methods with poor reproducibility.

Reference objects can contribute to validation and reproducibility of qMRI, and in this report, we reviewed the impact of reference objects on qMRI technique development and multi-site studies. Looking forward, the GE-NFL multi-site study and preclinical MRI

studies, as examples, clearly demonstrate the need for additional reference objects. In some cases, the NIST reference objects can be used as an initial source for the design of application-specific reference objects. A valid concern is that reference objects, even if there is a demonstrated need, will not be imaged on a regular basis if they are not easy-to-use and developed with software for automated analysis. Finally, when considering clinical use of qMRI, we need to know how often quality assurance or characterization of the systems are needed.

The clinical impact of differences in scanner software and hardware, as well as scanner performance over time, is not yet known; however, there is evidence that scanner variability can hinder the implementation of qMRI(21,93). With an appropriate reference object and a standard protocol, MRI systems can be characterized at regular intervals to develop appropriate quality assurance protocols that can mitigate the effects of scanner variability, such as has been done for specific studies(26,39,94). This recommendation is beyond the typical single-time site qualification. It can be challenging to understand the clinical utility of rapidly evolving technology coupled with instrument variability. Maximal scientific advancement is hindered in the absence of standards and reference objects.

Beyond quality assurance, to see the impact of quantitative MRI, we will need population level measurements of "normal" results. Bojorquez et al. present a compilation of the available T_1 and T_2 relaxation times at 3 T in a broad selection of tissues that to date is defined by the authors as *in vivo* measurements on normal subjects(95). Based on the range of relaxation times reported both within and across these studies, it is evident that more data collection in normal subjects is needed. Without such data characterizing normal results across a population, how do we know if a measure is in fact pathological? Without standardized data collection practices, including the use of reference objects, it will be hard, if not impossible, to compare data across centers, limiting the value of qMRI measurements.

In conclusion, to realize the value of qMRI for clinical diagnostics, treatment planning and outcomes assessment, the community needs to address several issues before or simultaneously with clinical validation, such as the technical verification of the MRI systems, acquisition techniques and analysis methods used to obtain them. Reference objects, both physical and computational, are a necessary component to the implementation of qMRI in a clinical setting, and a key conclusion from the workshop was that the widespread clinical adoption of qMRI requires ease-of-use, including easy reference object set-up, the smallest practical set of multi-parametric and possibly multi-modal reference objects for many techniques, and software for automatic analyses. Without addressing each of these components, the ability to implement qMRI in the clinic is likely to be limited.

REFERENCES

1. Kalmoe R, Mirowski E and Metzger GJ. Body Phantom with Prostate Mimic for Evaluation of Quantitative MRI. in ISMRM 26th Annual Meeting 2018 Paris, France (abstract 2762)
2. Metzger GJ, Kalavagunta C, Spilseth B, et al., Detection of Prostate Cancer: Quantitative Multiparametric MR Imaging Models Developed Using Registered Correlative Histopathology. Radiology, 2016 279(3):805–816. [PubMed: 26761720]

3. Boss MA, Chenevert TL, Waterton J, et al., Thermally-Stablized Isotropic Diffusion Phantom for Multisite Assessment of Apparent Diffusion Coefficient Reproducibility. Medical Physics, 2014 41(6):464.

4. Keenan KE PA, Wilmes LJ, Aliu SO, Jones EF, Li W, Kornak J, Newitt DC, Hylton NM, Variability and bias assessment in breast ADC measurement across multiple systems. Journal of Magnetic Resonance Imaging, 2016 44(4):846–855. [PubMed: 27008431]

5. Keenan KE WL, Aliu SO, Newitt DC, Jones EF, Boss MA, Stupic KF, Russek SE, Hylton NM., Design of a breast phantom for quantitative MRI. Journal of Magnetic Resonance Imaging, 2016 44(3):610–619. [PubMed: 26949897]

6. Price R, Allison J, Clarke G, Dennis M, Hendrick E, Keener C, Mastne J, Nessaiver M, Och J, Reeve D, MRI Quality Control Manual. 2015: American College of Radiology.

7. Russek SE, Boss MA, Jackson EF, et al. Characterization of NIST/ISMRM MRI System Phantom. in Proceedings of the 20th meeting of the International Society of Magnetic Resonance in Medicine, Melbourne, Australia 2012 (abstract

8. Delakis I, Moore EM, Leach MO and De Wilde JP, Developing a quality control protocol for diffusion imaging on a clinical MRI system. Phys Med Biol, 2004 49(8):1409–1422. [PubMed: 15152682]

9. Jackson EF, Bronskill MJ, Drost DJ, et al. Acceptance Testing and Quality Assurance Procedures for Magnetic Resonance Imaging Facilities. 2010.

10. Lerski RA and de Certaines JD, Performance assessment and quality control in MRI by Eurospin test objects and protocols. Magn Reson Imaging, 1993 11(6):817–833. [PubMed: 8371637]

11. Lerski RA, McRobbie DW, Straughan K, Walker PM, de Certaines JD and Bernard AM, Multi-center trial with protocols and prototype test objects for the assessment of MRI equipment. EEC Concerted Research Project. Magn Reson Imaging, 1988 6(2):201–214. [PubMed: 3374294]

12. Price RR, Axel L, Morgan T, et al., Quality assurance methods and phantoms for magnetic resonance imaging: Report of AAPM nuclear magnetic resonance Task Group No. 1. MEDICAL PHYSICS, 1990 17.

13. Wissmann L, Santelli C, Segars WP and Kozerke S, MRXCAT: Realistic numerical phantoms for cardiovascular magnetic resonance. J Cardiovasc Magn Reson, 2014 16:63. [PubMed: 25204441]

14. Barboriak DP. 2015 QIBA DRO for DCE MRI. Available: https://sites.duke.edu/dblab/qibacontent/ [10 June 2018].

15. Hernberg S, Lead Poisoning in a Historical Perspective. American Journal of Industrial Medicine, 2000 38:244–254. [PubMed: 10940962]

16. Bellinger DC, Bellinger AM, Childhood lead poisoning: the tortuous path from science to policy. J. Clinical Investigation, 2006 116(4):853–857.

17. Gilbert SG, Weiss B, A rationale for lowering the blood lead action level from 10 to 12 mg/dL. Neurotoxicology, 2006 27(5):693–701. [PubMed: 16889836]

18. Kessler LG, Barnhart HX, Buckler AJ, et al., The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Statistical Methods in Medical Research, 2015 24(1):9–26. [PubMed: 24919826]

19. Jain H, Kroening D, Sharygina N and Clarke EM, Word-level predicate-abstraction and refinement techniques for verifying RTL Verilog. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2008 27(2):366–379.

20. Clarke LP, Sriram RD and Schilling LB, Imaging as a biomarker - Standards for change measurements in therapy workshop summary. Academic Radiology, 2008 15(4):501–530. [PubMed: 18389935]

21. Keenan KE, Stupic KF, Boss MA, et al. Comparison of T1 measurement using ISMRM/NIST system phantom. in ISMRM 24th Annual Meeting 2016 Singapore (abstract 3290)

22. Pierpaoli C, Sarlls J, Nevo U, Basser PJ, Horkay F Polyvinylpyrrolidone (PVP) water solutions as isotropic phantoms for diffusion MRI studies. in Intl Soc Mag Reson Med. 2009 (abstract 1414)

23. Horkay F, Pierpaoli C and Basser PJ, inventorsPhantom for diffusion MRI Imaging. USA2012.

24. Freed M, de Zwart JA, Loud JT, et al., An anthropomorphic phantom for quantitative evaluation of breast MRI. Med Phys, 2011 38(2):743–753. [PubMed: 21452712]

25. Gunter JL, Britson PJ, Felmlee JP, et al. The ADNI phantom and analysis algorithm: a new and accurate tool to measure scanner performance in International Society of Magnetic Resonance in Medicine. 2007 Berlin, Germany (abstract

26. Gunter JL, Bernstein MA, Borowski BJ, et al., Measurement of MRI scanner performance with the ADNI phantom. Med Phys, 2009 36(6):2193–2205. [PubMed: 19610308]

27. Marquez D 11/2/201611/11/2016 2016.

28. Price R, Allison J, Clarke G, et al., 2015 MRI Quality Control Manual. 2015: American College of Radiology p.

29. Boss MA, Dienstfrey AM, Gimbutas Z, et al. Magnetic Resonance Imaging Biomarker Calibration Service: Proton Spin Relaxation Times. NIST Special Publication 250–97. National Institute of Standards and Technology, 2018 10.6028/NIST.SP250–97. 10.6028/NIST.SP.250-97

30. Jiang Y MD, Keenan KE, Stupic KF, Gulani V, Griswold MA, Repeatability of magnetic resonance fingerprinting T1 and T2 estimates assessed using the ISMRM/NIST MRI system phantom. Magnetic Resonance in Medicine, 2017 78(4):1452–1457. [PubMed: 27790751]

31. Chenevert TL, Galban CJ, Ivancevic MK, et al., Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. J Magn Reson Imaging, 2011 34(4):983–987. [PubMed: 21928310]

32. Malyarenko D, Galban CJ, Londy FJ, et al., Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging, 2013 37(5):1238–1246. [PubMed: 23023785]

33. Palacios EM, Martin AJ, Boss MA, et al., Toward Precision and Reproducibility of Diffusion Tensor Imaging: A Multicenter Diffusion Phantom and Traveling Volunteer Study. American Journal of Neuroradiology, 2017 38(3):537–545. [PubMed: 28007768]

34. Jezzard P and Balaban RS, Correction for geometric distortion in echo planar images from B0 field variations. Magnetic Resonance in Medicine, 1995 34(1):65–73. [PubMed: 7674900]

35. Bosca R, Ashton E, Zahlmann G and Jackson E. RSNA Quantitative Imaging Biomarker Alliance (QIBA) DCE-MRI Phantom: Goal, Design, and Initial Results in Radiological Society of North America 2012 Scientific Assembly and Annual Meeting. 2012 Chicago, IL (abstract

36. Gatehouse PD, Rolf MP, Bloch KM, et al., A multi-center inter-manufacturer study of the temporal stability of phase-contrast velocity mapping background offset errors. J Cardiovasc Magn Reson, 2012 14:72. [PubMed: 23083397]

37. Gatehouse PD, Rolf MP, Graves MJ, et al., Flow measurement by cardiovascular magnetic resonance: a multi-centre multi-vendor study of background phase offset errors that can compromise the accuracy of derived regurgitant or shunt flow measurements. J Cardiovasc Magn Reson, 2010 12:5. [PubMed: 20074359]

38. Rolf MP, Hofman MB, Gatehouse PD, et al., Sequence optimization to reduce velocity offsets in cardiovascular magnetic resonance volume flow quantification--a multi-vendor study. J Cardiovasc Magn Reson, 2011 13:18. [PubMed: 21388521]

39. Captur G, Gatehouse P, Keenan KE, et al., A medical device-grade T1 and ECV phantom for global T1 mapping quality assurance-the T1 Mapping and ECV Standardization in cardiovascular magnetic resonance (T1MES) program. J Cardiovasc Magn Reson, 2016 18(1):58. [PubMed: 27660042]

40. St Pierre TG, Clark PR, Chua-anusorn W, et al., Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. Blood, 2005 105(2):855–861. [PubMed: 15256427]

41. St Pierre TG, El-Beshlawy A, Elalfy M, et al., Multicenter validation of spin-density projection-assisted R2-MRI for the noninvasive measurement of liver iron concentration. Magn Reson Med, 2014 71(6):2215–2223. [PubMed: 23821350]

42. Hernando D, Sharma SD, Aliyari Ghasabeh M, et al., Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat-water phantom. Magn Reson Med, 2017 77(4):1516–1524. [PubMed: 27080068]

43. Hines CD, Yu H, Shimakawa A, McKenzie CA, Brittain JH and Reeder SB, T1 independent, T2* corrected MRI with accurate spectral modeling for quantification of fat: Validation in a fat-water-

SPIO phantom. Journal of Magnetic Resonance Imaging, 2009 30(5):1215–1222. [PubMed: 19856457]

44. Mashhood A, Railkar R, Yokoo T, et al., Reproducibility of hepatic fat fraction measurement by magnetic resonance imaging. Journal of Magnetic Resonance Imaging, 2013 37(6):1359–1370. [PubMed: 23172799]

45. NEMA Standards Publication MS 5–2009 Slice Thickness. 2009.

46. National Electrical Manufacturers Association. NEMA Standards Publication MS 1–2008 R2014 Determination of Signal-to-Noise Ratio SNR in Diagnostic MRI. 2014.

47. NEMA Standards Publication MS 3–2008 R2014 Image Uniformity. 2014.

48. International Electrotechnical Commission. Magnetic Resonance Equipment for Medical Imaging - Part 1: Determination of Essential Image Quality Parameters. cv. 2007.

49. International Electrotechnical Commission. IEC 60601–2–33 : Medical electrical equipment - Part 2–33: Particular requirements for the basic safety and essential performance of magnetic resonance equipment for medicel diagnosis. 2015.

50. 2018 American College of Radiology MRI Accreditation Program Requirements. Available: https://www.acraccreditation.org/modalities/mri [04/24].

51. Hines CD, Bley TA, Lindstrom MJ and Reeder SB, Repeatability of magnetic resonance elastography for quantification of hepatic stiffness. J Magn Reson Imaging, 2010 31(3):725–731. [PubMed: 20187219]

52. Shire NJ, Yin M, Chen J, et al., Test-retest repeatability of MR elastography for noninvasive liver fibrosis assessment in hepatitis C. J Magn Reson Imaging, 2011 34(4):947–955. [PubMed: 21751289]

53. Yasar TK, Wagner M, Bane O, et al., Interplatform reproducibility of liver and spleen stiffness measured with MR elastography. J Magn Reson Imaging, 2016 43(5):1064–1072. [PubMed: 26469708]

54. MR Elastography of the Liver. Committee QMEB. Quantitative Imaging Biomarkers Alliance, 2018 2018-05-02 DOI. http://qibawiki.rsna.org/index.php/Profiles

55. Messroghli DR, Moon JC, Ferreira VM, et al., Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). J Cardiovasc Magn Reson, 2017 19(1):75. [PubMed: 28992817]

56. Dyverfeldt P, Bissell M, Barker AJ, et al., 4D flow cardiovascular magnetic resonance consensus statement. J Cardiovasc Magn Reson, 2015 17:72. [PubMed: 26257141]

57. Moon JC, Messroghli DR, Kellman P, et al., Myocardial T1 mapping and extracellular volume quantification: a Society for Cardiovascular Magnetic Resonance (SCMR) and CMR Working Group of the European Society of Cardiology consensus statement. J Cardiovasc Magn Reson, 2013 15:92. [PubMed: 24124732]

58. Taouli B, Beer AJ, Chenevert T, et al., Diffusion-weighted imaging outside the brain: Consensus statement from an ISMRM-sponsored workshop. J Magn Reson Imaging, 2016 44(3):521–540. [PubMed: 26892827]

59. Reeder SB, Hu HH and Sirlin CB, Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. J Magn Reson Imaging, 2012 36(5):1011–1014. [PubMed: 22777847]

60. Hu HH, Bornert P, Hernando D, et al., ISMRM workshop on fat-water separation: insights, applications and progress in MRI. Magn Reson Med, 2012 68(2):378–388. [PubMed: 22693111]

61. Lingala SG, Sutton BP, Miquel ME and Nayak KS, Recommendations for real-time speech MRI. J Magn Reson Imaging, 2016 43(1):28–44. [PubMed: 26174802]

62. Inati SJ, Naegele JD, Zwart NR, et al., ISMRM Raw data format: A proposed standard for MRI raw datasets. Magn Reson Med, 2017 77(1):411–421. [PubMed: 26822475]

63. Alsop DC, Detre JA, Golay X, et al., Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med, 2015 73(1):102–116. [PubMed: 24715426]

64. ACR-ASNR-SPR Practice Parameter for the Performan of Intracranial Magnetic Resonance Perfusion Imaging. American College of Radiology, 2017 https://www.acr.org/-/media/ACR/Files/Practice-Parameters/MR-Perfusion.pdf

65. DePalma RG and Hoffman SW, Combat blast related traumatic brain injury (TBI): Decade of recognition; promise of progress. Behav Brain Res, 2018 340:102–105.

66. Cai H, Xu Z, Luo S, et al., Study on chemical fingerprinting of crude and processed Atractylodes macrocephala from different locations in Zhejiang province by reversed-phase high-performance liquid chromatography coupled with hierarchical cluster analysis. Pharmacogn Mag, 2012 8(32):300–307. [PubMed: 24082634]

67. Mac Donald CL, Barber J, Andre J, et al., 5-Year imaging sequelae of concussive blast injury and relation to early clinical outcome. Neuroimage Clin, 2017 14:371–378. [PubMed: 28243574]

68. Wilde EA, Provenzale JM, Taylor BA, et al., Assessment of quantitative magnetic resonance imaging metrics in the brain through the use of a novel phantom. Brain Inj, 2018 32(10):1266–1276. [PubMed: 30169993]

69. Bane O, Hectors SJ, Wagner M, et al., Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magn Reson Med, 2018 79(5):2564–2575. [PubMed: 28913930]

70. Partridge SC, Mullins CD, Kurland BF, et al., Apparent Diffusion Coefficient Values for Discriminating Benign and Malignant Breast MRI Lesions: Effects of Lesion Type and Size. American Journal of Roentgenology, 2010 194(6):1664–1673. [PubMed: 20489111]

71. Partridge SC, DeMartini WB, Kurland BF, Eby PR, White SW and Lehman CD, Differential diagnosis of mammographically and clinically occult breast lesions on diffusion-weighted MRI. Journal of Magnetic Resonance Imaging, 2010 31(3):562–570. [PubMed: 20187198]

72. Hambrock T, Somford DM, Huisman HJ, et al., Relationship between Apparent Diffusion Coefficients at 3.0-T MR Imaging and Gleason Grade in Peripheral Zone Prostate Cancer. Radiology, 2011 259(2):453–461. [PubMed: 21502392]

73. Provenzale JM, Mukundan S and Barboriak DP, Diffusion-weighted and Perfusion MR Imaging for Brain Tumor Characterization and Assessment of Treatment Response. Radiology, 2006 239(3):632–649. [PubMed: 16714455]

74. Chenevert TL and Ross BD, Diffusion imaging for therapy response assessment of brain tumor. Neuroimaging Clin N Am, 2009 19(4):559–571. [PubMed: 19959005]

75. Thoeny HC and Ross BD, Predicting and monitoring cancer treatment response with diffusion-weighted MRI. J Magn Reson Imaging, 2010 32(1):2–16. [PubMed: 20575076]

76. Partridge SC, Nissan N, Rahbar H, Kitsch AE and Sigmund EE, Diffusion-weighted breast MRI: Clinical applications and emerging techniques. J Magn Reson Imaging, 2017 45(2):337–355. [PubMed: 27690173]

77. Donati OF, Mazaheri Y, Afaq A, et al., Prostate cancer aggressiveness: assessment with whole-lesion histogram analysis of the apparent diffusion coefficient. Radiology, 2014 271(1):143–152. [PubMed: 24475824]

78. Diffusion-Weighted Magnetic Resonance Imaging (DWI). Committee QPDaFP-MB. Quantitative Imaging Biomarkers Alliance, 2017 20170427 DOI. http://qibawiki.rsna.org/index.php/Profiles

79. Simonelli J, Lee Y, Mikaiel S, et al. An MR-Compatible Stage for Respiratory Motion Emulation. in International Federation of Automatic Control. 2017 (abstract 6073–6078)

80. Mikaiel S, Simonelli J, Lee Y, et al. Hydrostatically Actuated MRI-Compatible Motion Platform for Dynamic MRI Research. in ISMRM 25th Annual Meeting 2017 Honolulu (abstract 5557)

81. Mikaiel S, Simonelli J, Lee Y, et al. Real-Time MRI-Guided Targeted Needle Placement During Motion using Rolling-Diaphragm Hydrostatic Actuators. in ISMRM 25th Annual Meeting 2017 Honolulu (abstract 736)

82. Yuh EL, Mukherjee P, Lingsma HF, et al., Magnetic resonance imaging improves 3-month outcome prediction in mild traumatic brain injury. Ann Neurol, 2013 73(2):224–235. [PubMed: 23224915]

83. Yuh EL, Cooper SR, Mukherjee P, et al., Diffusion tensor imaging for outcome prediction in mild traumatic brain injury: a TRACK-TBI study. J Neurotrauma, 2014 31(17):1457–1477. [PubMed: 24742275]

84. Shetty T, Nguyen JT, Cogsil T, et al., Clinical Findings in a Multicenter MRI Study of Mild TBI. Front Neurol, 2018 9:836. [PubMed: 30405511]

85. Li X, Bhushan C, Singanamalli A, et al. Single-subject voxel-based analysis for mTBI using Multi-shell Diffusion MRI. in Proceedings of the 23rd meeting of the Organization for Human Brain Mapping, Vancouver, Canada 2017 (abstract 3266)

86. Ravishankar H, Madhavan R, Mullick R, Shetty T, Marinelli L and Joel SE. Recursive feature elimination for biomarker discovery in resting-state functional connectivity. in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 (abstract 4071–4074)

87. Madhavan R, Joel SE, Mullick R, et al., Longitudinal Resting State Functional Connectivity Predicts Clinical Outcome in Mild Traumatic Brain Injury. J Neurotrauma, 2018, 10.1089/neu.2018.5739.

88. Zhang T, Masdeu J, Lebel R, et al., Longitudinal Analysis of Arterial Spin Labeling Perfusion in Mild Traumatic Brain Injury. Journal of Neurotrauma, 2016 33(13):A-31.

89. Jensen JH, Helpern JA, Ramani A, Lu H and Kaczynski K, Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn Reson Med, 2005 53(6):1432–1440. [PubMed: 15906300]

90. Yoshimaru E, Totenhagen J, Alexander GE and Trouard TP, Design, manufacture and anlysis of customized phantoms for enhanced quality control in small animal MRI systems. Magnetic Resonance in Medicine, 2013 71(2):880–884.

91. Doblas S, Almeida GS, Ble FX, et al., Apparent diffusion coefficient is highly reproducible on preclinical imaging systems: Evidence from a seven-center multivendor study. J Magn Reson Imaging, 2015 42(6):1759–1764. [PubMed: 26012876]

92. Lee YC, Fullerton GD, Baiu C, Lescrenier MG and Goins BA, Preclinical multimodality phantom design for quality assurance of tumor size measurement. BMC Med Phys, 2011 11:1. [PubMed: 21958653]

93. Stikov N, Boudreau M, Levesque IR, Tardif CL, Barral JK and Pike GB, On the accuracy of T1 mapping: searching for common ground. Magn Reson Med, 2015 73(2):514–522. [PubMed: 24578189]

94. Friedman L and Glover GH, Report on a multicenter fMRI quality assurance protocol. J Magn Reson Imaging, 2006 23(6):827–839. [PubMed: 16649196]

95. Bojorquez JZ, Bricq S, Acquitter C, Brunotte F, Walker PM and Lalande A, What are normal relaxation times of tissues at 3 T? Magn Reson Imaging, 2017 35:69–80.
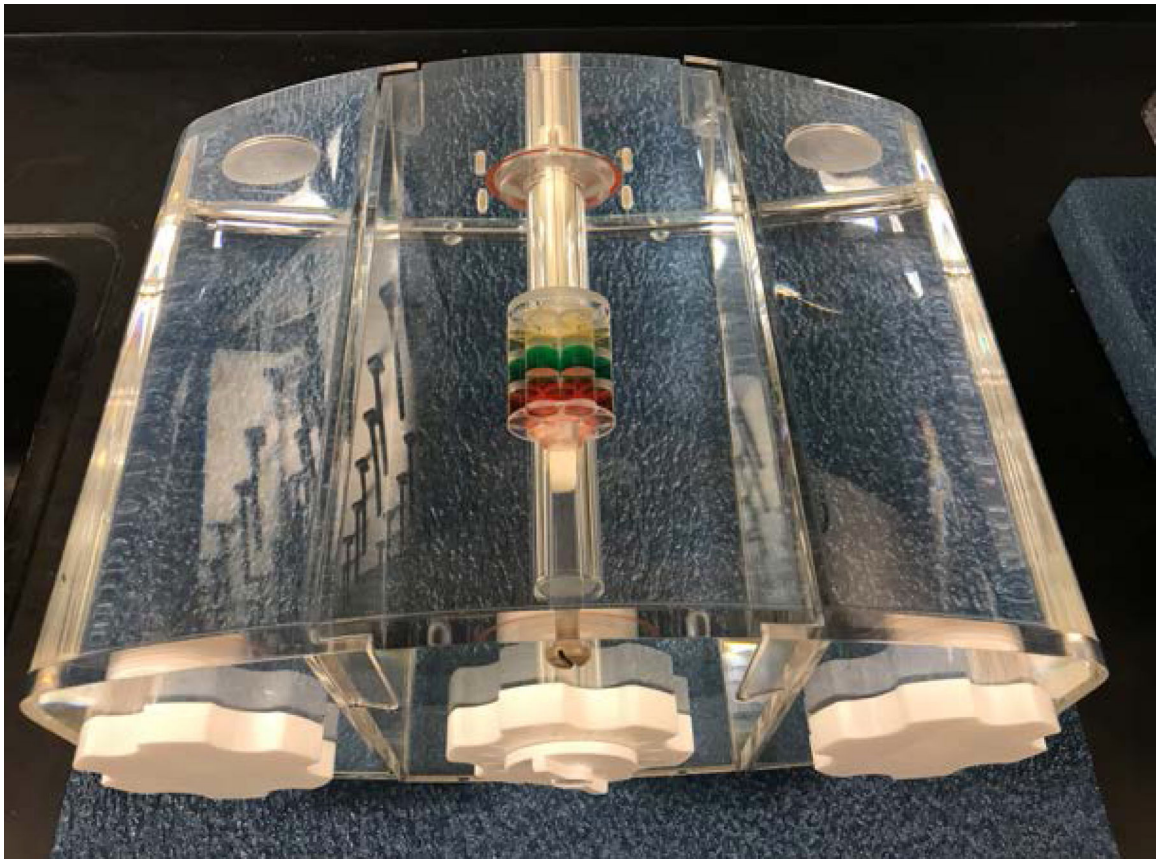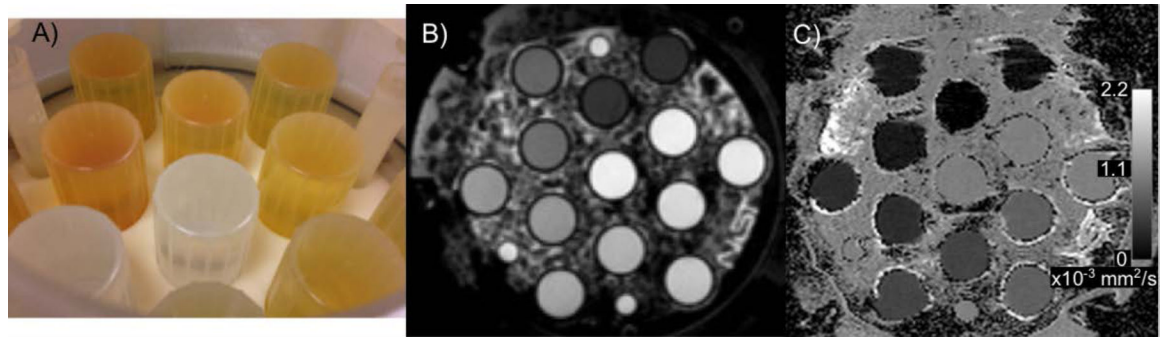
Figure 1:
Top-view of the UMN/HPD anthropomorphic, prostate, body-imaging reference object. This three-sectioned body phantom has a quantitative prostate mimic at its center (i.e. yellow, green and red structure) and can accommodate surface and endorectal coil configurations.
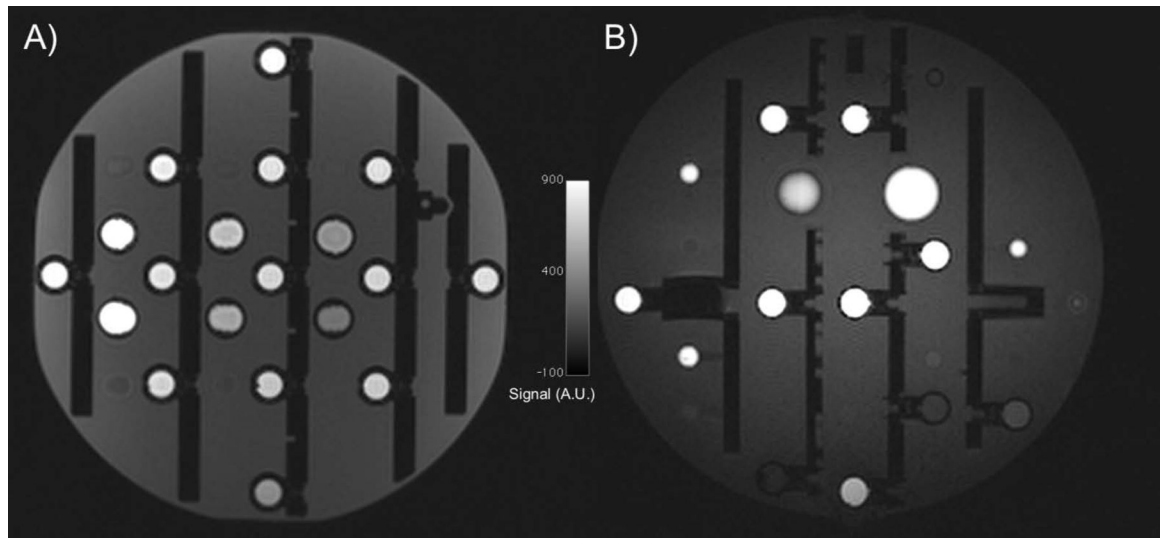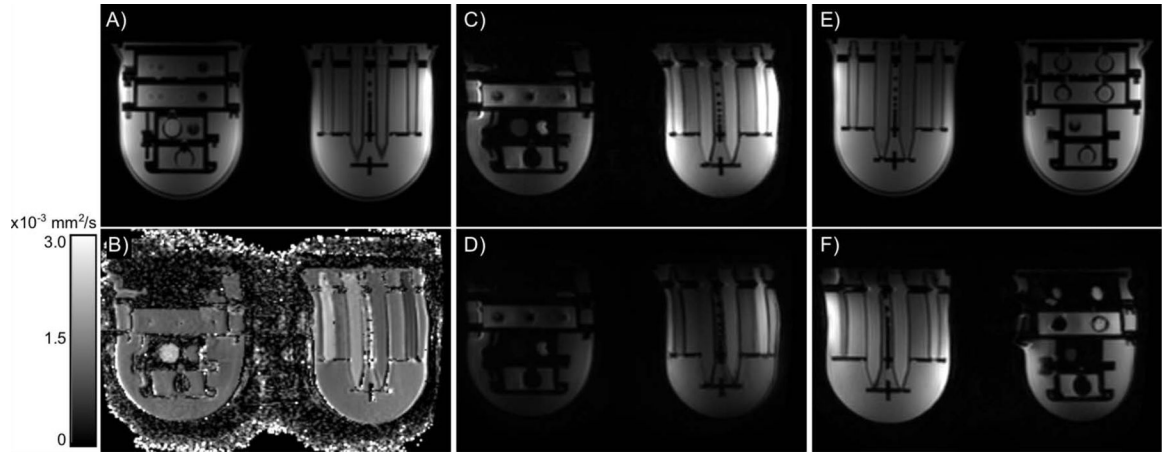
Figure 2:
Sagittal slices of a 3D T_1-weighted spoiled gradient echo sequence from the ISMRM/NIST system phantom (A) and the ADNI MagPhan phantom (B), which can be used to assess geometric distortions. An advantage of the ADNI MagPhan phantom in this orientation is the lack of symmetry.
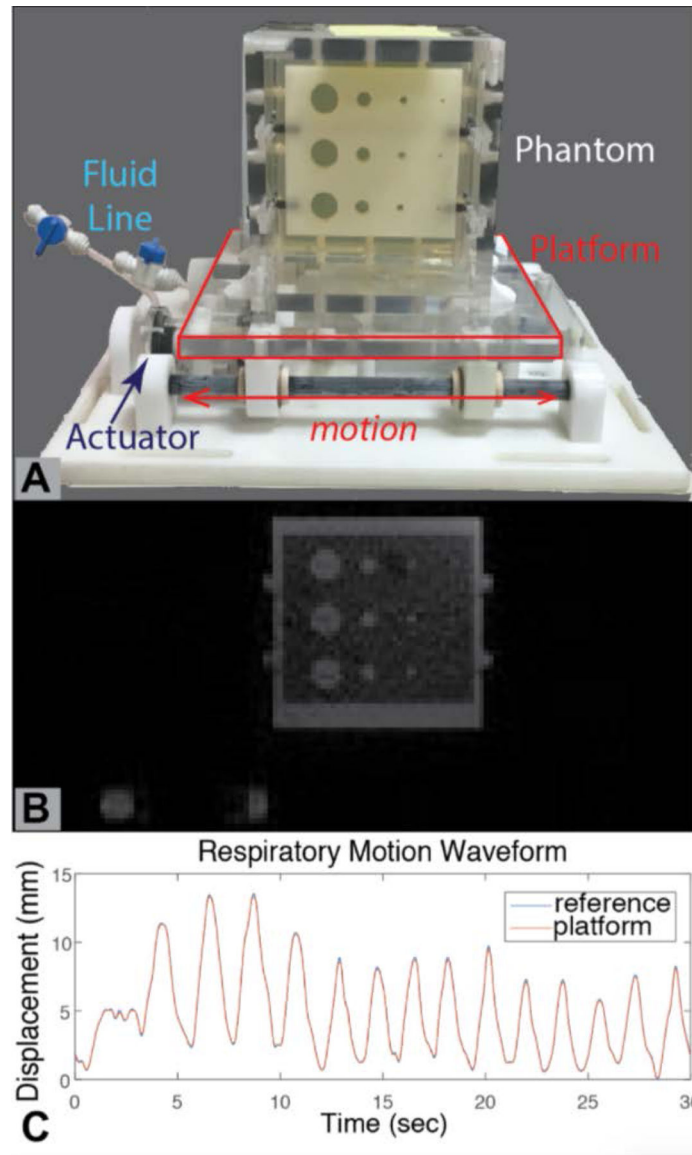
Figure 3:
NIST developed an isotropic diffusion phantom with support from NIH and RSNA-QIBA® based on the work of Pierpaoli *et al.* to identify an appropriate polymer, PVP, (22,23) and the ice-water phantom of Chenevert *et al* (31,32). A photo from the inside of the prototype diffusion phantom (A), a spin-echo image of the phantom filled with the ice water bath (B), and an apparent diffusion coefficient map with values in the phantom from approximately 0.2 to 1.1×10^{-3} mm^2/s (C).

Figure 4:
Axial images from a central slice of the UCSF/NIST prototype breast phantom. The T_1-weighted images (A, E) have no geometric distortion compared to the engineering designs, regardless of position in the coil. The EPI images (B-D, F) all demonstrate a geometric distortion in the right-left orientation: a stretch on the patient (and image) left and a shrinking on the patient (and image) right. In the ADC map (B), b-value = 0 image (C) and b-value = 800 s/mm^2 image (D), the object on patient (image) left is larger than the object on patient (image) right. Similarly, when the objects are reversed (E, F), the object on patient (image) left is wider in the b-value = 0 s/mm^2 image (F) compared to the T_1-weighted image (E).

Figure 5:
A programmable motion phantom for quantitative assessment of dynamic and interventional MRI (79). (A) The motion platform is actuated by a pair of master-slave hydrostatic actuators (slave actuator shown in photo). The slave and master actuators are connected by a fluid-filled line, with the master actuator positioned outside of the MRI scanner room and connected to a computer-controlled motor. The gel-filled phantom contains a plate with geometric features of known dimensions for assessment of dynamic MRI sequences. These features can also serve as targets for MRI-guided targeted device placement. The platform and phantom are constructed using plastic materials. (B) A representative real-time gradient echo MRI frame depicting the phantom and its geometric features. Note that the fluid (water) in the actuator is also visible. (C) By using a learning-based algorithm, the motion platform is programmed to accurately and repeatably generate motion as specified by a
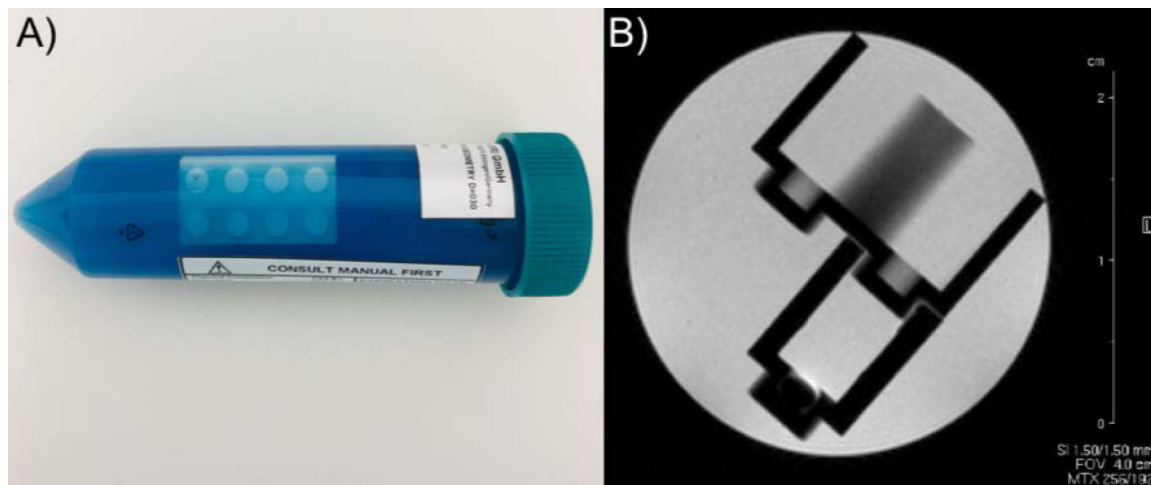
reference input waveform. In this example, the reference input was an actual respiratory waveform pre-recorded from a human subject.

Figure 6:
Photo of a manufacturer-supplied reference object for a 9.4 T pre-clinical imaging system
from August 2018 (A) and an acceptance image using a similar manufacturer-supplied
reference object for a 4.7 T pre-clinical imaging system from June 2005 (B). The pre-
clinical imaging reference objects are simply conical 50-cc tubes filled with fluids for
contrast (no relaxation time measurements) and small, interlocking, plastic, toy "bricks" for
volumetric measurements. There was no significant improvement in reference object design
for over a decade.

Table 1:

Organizations working towards standards in quantitative MRI who participated in the 2014 and 2017 NIST workshops

Acronym	Organization
AAPM	American Association of Physicists in Medicine
ACR	American College of Radiology
ACRIN	American College of Radiology Imaging Network
DoD	Department of Defense
FDA	Food and Drug Administration
IAC	Intersocietal Accreditation Commission
ISMRM	International Society of Magnetic Resonance in Medicine
MITA/NEMA	Medical Imaging and Technology Alliance/ National Electrical Manufacturers Association
NIH	National Institutes of Health
NCI	National Cancer Institute
NIST	National Institute of Standards and Technology
PhRMA	Pharmaceutical Research and Manufacturers of America
RSNA	Radiological Society of North America
QIBA	Quantitative Imaging Biomarkers Alliance
QIN	Quantitative Imaging Network (under the NCI Cancer Imaging Program)
VA	Department of Veteran's Affairs

Table 2:

Specific reference object and measurement needs identified at the 2017 NIST workshop[1]

Low-cost/home-made phantom recipes to help get sites started on standardized quantitative MRI
Make reference standards more biologic (e.g., including magnetization transfer, microstructure, etc)
Improve ability of reference objects to account for spatial variations in B_1 and B_0 (e.g., distribute samples for given parameter throughout phantom or vary all parameters within samples)
Consider more advanced contrast media (e.g., T_1 simultaneous with T_2 or multi-exponential diffusion with realistic T_1 and T_2)
Next generation phantom to include physiologically relevant values of T_1, T_2 and ADC in each voxel
Phantom sensitive to magnetization transfer (MT) effects, both to measure MT and to optimize fast T_2 mapping techniques so that T_2 values are not affected by MT
Consider next generation of diffusion phantoms (DTI applications)
Phantom with compartments that each simultaneously have diffusion-T_1-T_2-R_2^*-PDFF (proton density fat fraction) characteristics and span relevant physiological values
R_2^*/fat fraction phantom
Dynamic phantom (e.g., physiological motion)
Pre-clinical (animal) phantom
Multi-modality (MR-PET) phantom
Quantitative flow phantom
Musculoskeletal (MSK) phantoms (by considering the coil shape; can start with knee coil); T_{1rho} measurement; the tissue relaxation properties of long T_1 and short T_2; temperature monitoring and control etc.

[1] The order of the list should not be interpreted as a ranking of priority.

[*] Certain commercial instruments and software are identified to specify the experimental study adequately. This does not imply endorsement by NIST or that the instruments and software are the best available for the purpose.

[*] Certain commercial instruments and software are identified to specify the experimental study adequately. This does not imply endorsement by NIST or that the instruments and software are the best available for the purpose.