



Metrology for MRI: the field you've never heard of

Matt G. Hall¹ · Matt Cashmore¹ · Hyo-Min Cho² · Bernd Ittermann³ · Kathryn E. Keenan⁴ · Christoph Kolbitsch³ · Changwoo Lee² · Chengwei Li⁵ · Asante Ntata¹ · Katie Obee¹ · Zhang Pu⁵ · Stephen E. Russek⁴ · Karl F. Stupic⁴ · Lukas Winter³ · Luca Zilberti⁶ · Michael Steckner⁷

Received: 11 November 2024 / Revised: 31 January 2025 / Accepted: 14 February 2025 / Published online: 19 March 2025
 © The Author(s) 2025

Abstract

Quantitative MRI has been an active area of research for decades and has produced a huge range of approaches with enormous potential for patient benefit. In many cases, however, there are challenges with reproducibility which have hampered clinical translation. Quantitative MRI is a form of measurement and like any other form of measurement it requires a supporting metrological framework to be fully consistent and compatible with the international system of units. This means not just expressing results in terms of seconds, meters, etc., but demonstrating consistency to their internationally recognized definitions. Such a framework for MRI is not yet complete, but a considerable amount of work has been done internationally towards building one. This article describes the current state of the art for MRI metrology, including a detailed description of metrological principles and how they are relevant to fully quantitative MRI. It also undertakes a gap analysis of where we are versus where we need to be to support reproducibility in MRI. It focusses particularly on the role and activities of national measurement institutes across the globe, illustrating the genuinely international and collaborative nature of the field.

Keywords MRI · Quantitative MRI · Metrology · Reproducibility · Traceability · Uncertainty

Abbreviations

AAPM	American association of physicists in medicine
ACR	American college of radiology
ADC	Apparent diffusion coefficient
ADNI	Alzheimer's disease neuroimaging initiative
AI	Artificial intelligence
APMP	Asia-Pacific metrology programme

BIPM	Bureau International des Poids et Mesures (the International Bureau of Weights and Measures)
BCE	Before the common era
BSI	British Standards Institute
CEI	Italian Electrotechnical Commission
DCE	Dynamic contrast-enhanced
DL	Deep learning
DUT	Device under test
EIBALL	European Imaging Biomarkers Alliance
EMPT	Electromagnetic tissue properties
EPT	Electrical properties tomography
FDA	Food and Drug Administration (US medical regulator)
GUM	Guide to the expression of uncertainty in measurements
IAEA	International Atomic Energy Agency
IEC	International Electrotechnical Commission
IEEE	The Institute for electrical and electronics engineers
IPEM	Institute of physics and engineering in medicine
IQ	Image quality
IR	Inversion recovery

✉ Matt G. Hall
matt.hall@npl.co.uk

- ¹ National Physical Laboratory, Teddington, UK
- ² Korea Research Institute of Standards and Science, Daejeon, Republic of Korea
- ³ Physikalisch-Technische Bundesanstalt, Berlin, Germany
- ⁴ National Institute of Standards and Technology, Boulder, CO, USA
- ⁵ National Institute of Measurement, Beijing, People's Republic of China
- ⁶ Istituto Nazionale Di Ricerca Metrologica, Turin, Italy
- ⁷ MKS Consulting, Beachwood, OH, USA

ISMRM	International Society for Magnetic Resonance in Medicine
INRIM	Istituto Nazionale di Ricerca Metrologica (Italian NMI)
ISO	International organisation for standardization
J_QIBA	Japan Quantitative Imaging Biomarkers Alliance
KRISS	Korea Research Institute of Standards and Science (South Korean NMI)
KSA	Korean Standards Association
MDR	Medical device regulation
MAFLD	Metabolic dysfunction-associated fatty liver disease
MASLD	Metabolic associated steatotic liver disease
MMFG	Focus group of medical metrology
MOMA	Modular mapping
MR	Magnetic resonance
MRI	Magnetic resonance imaging
NEMA	National Electrical Manufacturers Association
NIM	National Institute of Measurement (Chinese NMI)
NIST	National Institute of Standards and Technology (US NMI)
NMI	National Metrology Institute
NMPA	National Medical Products Association (China's medical regulator)
NPL	National Physical Laboratory (UK NMI)
NMR	Nuclear magnetic resonance
PDFF	Proton density fat fraction
PTB	Physikalisch-Technische Bundesanstalt (German NMI)
QA	Quality assurance
QC	Quality control
qMRI	Quantitative magnetic resonance imaging
QUIERO	Quantitative imaging enables reproducible outcomes
RF	Radiofrequency
RSNA	Radiological society of North America
RSNA-QIBA	RSNA Quantitative Imaging Biomarkers Alliance
SAR	Specific absorption rate
SI	Système International
SNR	Signal to noise ratio
TIMES	T1-mapping and extracellular volume standardisation
UKAS	UK accreditation service
VIM	International Vocabulary of Metrology

Introduction

Quantitative methods for magnetic resonance imaging (MRI) have been in existence since the technology's invention [1]. In the decades since, enormous strides have been and continue to be made across a range of MR-based imaging methods and post-processing approaches [2]. From beginnings in mapping magnetic relaxivities, MRI is now routinely used to measure quantities as diverse as diffusion [3], flow [4], magnetic susceptibility [5], fat fraction [6], iron content [7], and many others. Quantitative MRI (qMRI) is also used to measure volume, shape, and position [8] of tissue features in a wealth of applications. Whilst these do not require quantitative contrast, they are nonetheless measurements derived from images although we will not define them as quantitative MRI for the purposes of this article, which concentrates on quantitative MRI contrast methods.

Despite this, MRI as used in the clinic is typically not quantitative. The overwhelming majority of MRI procedures performed in hospitals are aimed at generating images to be interpreted visually by human experts to provide information for the diagnostic process. Rather than mapping the value of a particular physical parameter across an organ or region, clinical images typically make use of relative contrast, and spatial inference is subjective and relative to other, nearby tissue features [9]. Quantitative MRI is used in individual diagnostic applications such as characterization of liver iron in haemochromatosis [10], but typical clinical usage is more often semi-quantitative such as when a quantitative method is applied in a non-quantitative way (such as the use of diffusion-weighted MRI to diagnose ischemic stroke [11]), or where imaging outputs themselves are not fully quantitative such as dynamic contrast-enhanced (DCE) imaging [12]. Here effective parameters are extracted from images which are not fully reproducible across scanners and sessions.

What is the reason for this apparent disparity? It is certainly not a lack of potential clinical utility. Publications and reviews of qMRI approaches are very commonly motivated by application to pathology, whether it be diagnosis, staging, treatment planning, or assessing impact [13], and whilst clinical applications are not the only possible uses of qMRI approaches (clinical trials, for instance, are also a key area of application), they certainly represent a very major use case. Applications in clinical trials are similarly not widespread, and larger scale research studies such as the UK Biobank [14] mostly employ routine clinical approaches.

One, at least partial, answer to this question is reproducibility. For a quantitative method to be genuinely useful, it needs to be reliable. Blood tests and other chemical

assays are widely used precisely because the answers can be trusted—a clinician knows that a blood sample from a patient will be processed to a high standard and that error rates are understood and minimized [15]. This is also true in the delivery of therapeutics—doses either of pharmaceutical agents or of radiation are highly standardized and extremely reliable. Clinicians can trust that the numbers are what they say they are, and strategies calibrated accordingly.

Such reliability and trustworthiness are not an accident or an inherent property of the technologies. This reproducibility is supported by a framework of measurement science which assesses uncertainty, provides consistent calibration, and provides both guidance for practitioners and support for the standards which the field relies upon. This framework is metrology, the science of measurement.

Metrology ensures that wherever you are in the world a meter is a metre, a gram is a gramme, and a second is a Sekunde. Metrology in its modern form has existed for over a century,¹ it underpins measurements in labs, hospitals, factories, and courtrooms across the world [17] [18, 19], but for various historical reasons metrology is a relatively recent entrant into the field of MRI. This review will describe what metrology is, why it is important, and what work has already been done. It will also take a more forward-looking view on where MRI metrology needs to be to properly support the development of qMRI, its translation into clinical and other applications, and to maintain quality in its day-to-day use.

Metrology is inextricably linked to the development of standards and for this reason we will also include a review of the current standards landscape and cover some new directions. Unhelpfully, the word “standard” has more than one definition in metrological literature (at least in the English language subset of it). Depending on context, it can mean a document describing how or how well a particular procedure should perform or be performed (as in, for example, an International Electrotechnical Commission (IEC) standard [20]). “Standard” also refers to a measurement device or process which is used as a reference to calibrate another measurement device or process [21]. This review discusses both kinds of standards, but for clarity we will refer to the former as “documentary standards” or “written standards” and the latter as “reference standards” or “metrological standards”. Please note that this is not the case in the rest of the metrological literature, however.

¹ Arguably the birth of modern metrology began in France towards the end of the French Revolution. Seeking an end to the proliferation of mismatching units, the Academy of Sciences tasked a committee to define a new fundamental unit of length; the Meridian. This eventually led to the admission of the infamous Platinum Metre, as well as the Platinum Kilogramme, to be deposited within the French National Archives in July 1799 [16].

The remainder of this document is structured as follows. First, we will describe metrology and how it works, with examples and focus on global metrological structures and mechanisms, and how these are relevant to MRI. We will then briefly review metrology in other fields of medical physics and imaging, before moving on to qMRI and its needs from a metrological perspective. We will then review the current metrological landscape for MRI, including both quantitative applications and more traditional approaches. Finally, we will take a more forward-looking approach and make a case for an improved metrological framework which better supports the qMRI community.

What is metrology?

Metrology is the science of measurement [22]. Fundamentally metrology is about ensuring consistency and comparability in measurements made in different places, at different times, and by different methods. One of its key contributions is ensuring consistency in units. Through the activities of National Measurement Institutes (NMIs), manufacturers, researchers, practitioners, engineers, the legal profession, and many others can have confidence that no matter where or how a measurement is made Metrology ensures that the *Système international d'unités* (SI) system of units is consistent and reliable [22]. Key reference texts for metrology are the *Guide to the Expression of Uncertainty in Measurement* (“the GUM”) [23] and the *International Vocabulary of Metrology* (“the VIM”) [22] and we will reference these extensively in this work.

Historically, the role of metrology in the practice of MRI has been limited to a few areas such as implant safety testing or the manufacturing process itself [24–26]. MR images as they are used in the clinic are typically not measurements per se. In the research space, however, things are very different. qMRI has been a reality for decades—MRI is capable of measurements of a wide variety of quantities, from relaxivities to diffusivities to susceptibilities, and as noted above even conventional clinical MR images [1] can be used to measure volumes, positions, and shapes of organs or tissue features. In this review we concentrate on the first of these forms of measurement—the formation of parameter maps. This is partly because this is closer to the accepted definition of “quantitative MRI” as it is typically used, and because this is where the bulk of recent metrological development has happened.

Quantitative MRI, of any form, is measurement. Through the application of a specific pulse sequence, readout, and analysis pipeline the quantities derived in either quantitative maps or via spatially extended inference are physical

parameters which, with perhaps some exceptions,² can be said to exist independently of the scan process used to measure it. This means that the tools and frameworks of metrology are available to us. Metrology offers the opportunity not just to characterize reproducibility of results, but to understand the limitations of a given dataset.

Metrology's link to standardisation and documentary standards come from providing evidence of consistency and underpinning reproducibility. Metrology also links to best practice and measurement performance standards which are codified in national and international standards. Documentary standards are extremely important, not least because they are the mechanism by which the manufacturers of commercial products demonstrate regulatory compliance. Standards need to reflect the best-available knowledge of a technique and enable innovation in a particular technology. Standards also require monitoring to ensure that they effectively support new and emerging techniques without hindering the use of existing ones. In the case of measurements, this necessitates that the existing metrological infrastructure is in place to support quantitative measurements, without which there can be no appropriate basis on which to perform any validation.

A key feature of metrology is that it allows the quantitative comparison of measurements of the same quantity made in different ways. This is crucial for assessing the performance of a measurement process, since it allows measurements from the new process to be compared to other existing ones we already have confidence in. One measurement process can be used to calibrate another, despite both being imperfect to one extent or another. We might, for example, calibrate a practical, on-the-ground measurement approach to another which is more precise but less convenient or flexible. This is known as reference metrology, and leads to the idea of a metrological standard—physical objects or processes which have properties characterized by an independent reference process [22]. The process by which we relate the on-the-ground, day-to-day measurement to the most precise method we have available for the realization of the particular unit is called traceability, and this is discussed in detail later in this review.

² An argument might be made that relaxivities such as T1 and T2 can only be said to exist for samples placed in a strong magnetic field and as such are not meaningful unless the object is in a scanner. We might further argue that these are phenomenological parameters from a coarse-grained description of underlying quantum processes. The counter to this would be that the relaxivities we measure are a consequence of microscopic molecular motion, which is always occurring, and that nuclear spins will exhibit the same behaviour in the same nominal field regardless of its source (or the manufacturer of its source). The fact that the definition of the parameters comes from the Bloch equations rather than a quantum–mechanical description of spin–spin or spin–lattice relaxation is simply a choice of model and therefore, metrologically, just another source of bias and uncertainty.

Reference standards can in principle themselves be calibrated to other, even more precise standards, leading eventually to the definition of SI units. This hierarchical approach allows very large numbers of measurement devices to be calibrated to common standards internationally. This system also avoids the need for a hypothetical perfect ground truth whilst still allowing consistency, something we will describe in detail later. Providing these references and the link to the SI system is the role of NMIs and we will also describe their role in MRI metrology, including the current state of the art and some current lines of development.

If, as a practitioner of MRI, these concepts seem unfamiliar then you are certainly not alone. Aside from applications in device safety, metrology for MRI is largely unknown outside of a relatively small community, but as interest in quantitative imaging and its translation into clinical use grows, awareness of the need for metrology is growing, as evidenced by recent activity at the meeting of the International Society for Magnetic Resonance in Medicine (ISMRM) [27].

Why we need it

Consistency of measurement is critical for a number of reasons. The first is to ensure and demonstrate reproducibility and underpin the evidence behind new research claims. It is essential that new data can be compared against existing datasets and that experimental evidence from one lab is directly comparable to similar data from a different lab. Without this, science cannot function. Metrology also supports the spread of best practice in measurement. By quantifying measurement performance, it is possible to standardize what is possible in a particular context, and also to make recommendations on which particular method in practice leads to optimal performance.

In MRI, new techniques are frequently developed by researchers and then translated into clinical use by commercial manufacturers who make approaches practical and economical and guide them through regulatory approval. These are critical steps in achieving patient benefits. Consistency and comparability of measurement are essential in supporting effective translation and use. Any new approach that estimates a physical quantity needs to be as reproducible across devices so that the measured value and any changes can be assessed for significance. This is as true of a cardiac T1 value as it would be of a biochemical assay, a temperature, or the orbital trajectory of a satellite. Reproducibility means that differences between patients can be quantified, and individuals can be stratified into groups. Reproducibility is also essential for calculating effect size in a large-scale study such as a clinical trial.

In each of these cases, it is also necessary to compare measurements directly. This means understanding not just the differences between the measured values but also the

significance of those differences. Significance of difference is an inherently statistical question and requires an understanding of the uncertainty in each measurement. In strict metrological terms, the difference between measurements is a statistical statement of significance. Measurements are fully described by distributions [19] and difference or similarity is stated in terms of statistical significance. Two measurements can only be meaningfully compared if the measurement uncertainties are understood and quantified for both. Essentially, we are asking if two measured values are likely to have drawn from the same distribution. No two measured values will ever be exactly the same, but by evaluating the uncertainties associated with them, we gain information about the probability distribution a measurement was sampled from and thereby ask how significant the difference is.

Measurement uncertainty is also important when comparing a measured value to a known threshold e.g., a known safe limit for dosage or exposure. A quantified uncertainty provides important contextual information. A measured value may well be below a defined threshold, but if the range of values defined by the uncertainty breaches it, then additional steps should be taken.

Metrology treats measurement performance in terms of uncertainty and bias [28]. These are statistically and quantitatively well-defined terms which allow a fully quantitative approach to measurement performance evaluation. Metrological terminology also tends to avoid using the word “error” (except in certain, well-defined contexts [22]) and the word “accuracy”. The reason for this is that the concepts of both accuracy and error³ involve comparing a measured value to an underlying “ground truth”, which is known with infinite precision. Since no measurement procedure can ever produce a value known with infinite precision and would require knowledge of every effect capable of affecting the measurement, a reference ground truth of this kind can never be known and as a result neither the accuracy nor the error of a measurement can ever be fully quantified. However, the uncertainty and the observed bias from a reference value can both be quantified. Rather, metrology seeks to establish a description of a measurement phenomenon where we present a numerical result along with a statistical description of the range of values in which we believe the “true” result resides. Uncertainty is defined as the breadth of the distribution describing a set of repeated measurements of the same quantity. It can be thought of as the limit of information we have about the underlying value. Bias is defined as a systematic shift away from a reference value. Systematic bias is

often easier to deal with than uncertainty, since a systematic shift can often be calibrated away once the measurement process is well characterized.

With enough information about measurement uncertainty, we can formulate the comparison of two measurements within a formal hypothesis testing framework and obtain statistical information (such as standard deviations, significance tests, etc.) to help guide decisions and interpretation [23]. A formal statistical comparison framework for measurements also allows measurement equipment to be calibrated. By performing measurements using one approach on an object which has already been characterized with low uncertainty using a different approach, we can obtain information not just about the uncertainties associated with the measurement process, but also about any systematic biases it exhibits. As we will see in the next section, calibration is also critical in ensuring consistency of units.

A fair criticism at this point would be that comparing a measurement to a reference value seems to be just pushing the same problem onto the next measurement—if we need a reference value to assess measurement performance, how can we assess the measurement performance of the reference, and so on? The answer is *by focussing on consistency in units*. For example, two measurement processes with quantified uncertainties that are calibrated to the same system of units will produce comparable results. It is, therefore, critical that all measurement processes agree on the definition of the units.

The fact that comparability between measurements can be almost taken for granted in many fields is testament to the metrological frameworks that underpin them. A 10 mm bolt will fit a 10 mm nut, and if it doesn’t, the user immediately knows that there is a flaw in the manufacturing process of one or both, damaging the reputation of the manufacturer. Indeed, in the field of engineering the principle of tolerancing is born out of an understanding of inherent variation, both in the manufacturing process and the measurement processes used to monitor it. This is especially critical where safety is important. A pharmaceutical manufactured with a certain stated dose must be trustworthy for the physician prescribing it or patients could be harmed. Radiation delivered to a patient undergoing radiotherapy is carefully calibrated and directed to maximize the dose to a tumor while minimizing the dose to healthy surrounding tissue. In each case, uncertainties are carefully characterized and tightly controlled, and equipment regularly audited to ensure its performance remains within accepted guidelines [29].

MRI has been used to make measurements in research studies for decades including relaxivities, diffusivities, chemical compositions, flow fields and susceptibilities. The list is long [1]. In each case an MR scanner is used to apply a pulse sequence which sensitizes received signals to a particular physical mechanism, images are formed

³ Unlike accuracy and error, precision is a fully quantitative term. Precision refers to the spread of values if a single measurement is repeated many times and as such can in principle be known without reference to a hypothetical perfect ground truth.

and processed, models are fitted, and maps of parameters generated.

To fully and quantitatively assess the performance of a qMRI technique (of whatever kind) and to ensure that the results of one study are fully comparable to those of another, we need the same kind of metrological framework as has been developed in manufacturing, pharmaceuticals, or radiation dosimetry, but one that has been specifically designed for MRI. Such a framework would then underpin not just reproducibility in research work, but other ongoing standardisation and reproducibility efforts and thereby enable the widespread use of qMRI in routine healthcare and clinical trials.

Traceability and the role of national metrology institutes

Units are the scales we use to compare quantities and outputs. A consistent system of units is absolutely critical to the modern world since it enables consistency, interoperability, trade, and ultimately trust in products and services. The need for consistency in units has been recognized since antiquity and led historically to establishing bodies with responsibility for maintaining it. The earliest known well-defined unit was the Royal Egyptian Cubit, which was defined by a block of black granite established during the building of the great Khufu pyramid in approximately 9000 BCE⁴ [30]. This was a standard for length based on the distance between the pharaoh's elbow and forearm, plus the width of the palm. Other examples from history include the regulation of lengths and weights during Qin dynasty China in 221 BCE [31], the Platinum metre from revolutionary France [16], and reference length standards at the Greenwich observatory in London [32]; all used as references in industry and trade to ensure consistency. Similar needs were experienced in the US after independence, and as early as 1807 George Washington petitioned congress “to fix the standards of weights and measures” to achieve consistency across state boundaries eventually leading to a set of standards established by George R Hassler [30].

The modern era of metrology began with the establishment of the SI in the twentieth century. This established a set of base units (the second, meter, kilogram, kelvin, ampere, mole,⁵ and candela) with internationally agreed-upon definitions and overseen by the Bureau International des Poids et Mesures (BIPM). These definitions are linked to practical applications of measurement via a set of reference standards

maintained by national-level bodies called National Metrology Institutes (NMIs). NMIs provide the link between measurements made in practical, on the ground activities and the internationally accepted definitions of units.

Measurements require units.⁶ For measurements to be fully consistent and comparable, units need to be consistent. The kelvin (or celsius) degree as measured on a thermometer should be consistent with the definition of the SI kelvin. Without a common definition, kelvin or celsius temperatures used in one place, time, or study will be inconsistent with those in another. This leads to the concept of traceability. Traceability means that any given measurement device is connected, via a suitable, unbroken chain of calibrations, to the internationally accepted definition of the relevant SI unit or units. A measuring device which is not traceable to an SI unit is of limited utility.

The SI system defines the seven base units. Since 2019, these have been defined by fixing the numerical values of fundamental constants (such as Plank's constant, the speed of light, the frequency of a well-known transition in the electronic structure of Cesium, and a few others) to internationally agreed-upon values [33]. These definitions are extremely robust but not necessarily practical for wide deployment, so the BIPM also defines “embodiments” of units which are a set of experimental procedures to establish their values. From these the multitude of other units in modern life such as density, luminance, magnetic field strength, etc. all originate. These are called Derived Units.

The SI system is connected to other applications of measurement via NMIs. The role of an NMI is two-fold: to develop and maintain what are known as primary standards, which are devices which embody the relevant SI unit for practical purposes, and to offer measurement and calibration services which allow other devices to be calibrated to them, providing traceability. In the early years of modern metrology, NMIs typically did only this, and their activities were limited to the SI units alone. Modern NMIs now typically offer a broad range of metrological standards and services to provide traceability in as wide a range of specialties as possible. SI traceable standards and services exist for radiation dose [34] and gas concentration [35], and many NMIs have activities in mathematical and data-oriented applications such as cryptography [36] and AI-based inference [37]. Many NMIs are also involved with Legal Metrology, which covers how one applies regulatory and statutory mechanisms to the enforcement of metrology and metrological guarantees to the legal profession when needed.

⁴ Before the common era.

⁵ The mole was a little late to the party. When the original SI was introduced globally in 1960 there were only 6 units, with the mole being a brought into the family at the 14th Conférence générale des poids et mesures (GCPM) in 1961 [33].

⁶ We could argue that a percentage or a proportion is unitless but still quantitative, and this would not be unreasonable. To assess the performance of an estimate of proportion, however, we would still require unitful estimates of the various constituents, e.g., total volume of substance of interest and total volume of object. Often our references may be unitful even if the outputs are not.

Typically, each industrialized nation will have its own NMI. Examples include the US National Institute of Standards and Technology (NIST), Physikalisch-Technische Bundesanstalt (PTB) in Germany, Istituto Nazionale di Ricerca Metrologica (INRIM) in Italy, China's National Institute of Measurement (NIM),⁷ the Korea Research Institute of Standards and Science (KRISS), and the UK's National Physical Laboratory (NPL). This is also the list of NMIs globally who (as of 2025) have teams and activities specifically in MRI. Part of maintaining the consistency of the SI system is performing comparisons between measurement capability and primary standards. It is important to benchmark different capabilities against each other, and to identify and remediate any inconsistencies. These are known as International Inter-comparisons and NMIs typically perform them regularly, registering the results with the BIPM. BIPM is not a measurement institute and does not maintain standards. It exists as the highest level of metrological organisation, allowing NMIs to compare capabilities and develop new activities.

Whilst it is possible for any given measurement device to be calibrated directly to a primary standard, this is typically not how traceability works in practice. A single, national-level institution does not have the capacity to calibrate every measurement device being used. Instead, it is more common for NMIs to calibrate other standards, which are then used to calibrate others and so on. Standards calibrated directly to a primary standard are known (unsurprisingly) as secondary standards, which in turn are used to calibrate tertiary standards and so on. Depending on the scale of the measurement infrastructure, secondary standards may be used by calibration labs, who then offer calibration services to organizations performing measurements in practical applications, often calibrated via working standards maintained in-house. This is referred to as the traceability pyramid, as can be seen in Fig. 1. Each layer of calibration increases capacity in the system since each reference standard can be used to perform many calibrations but leads to an increase in measurement uncertainty.

Organizations engaged in calibration for SI traceability are typically accredited to do so by a national-level body (e.g., the UK Accreditation Service (UKAS)) which audits them for compliance with documentary standards such as

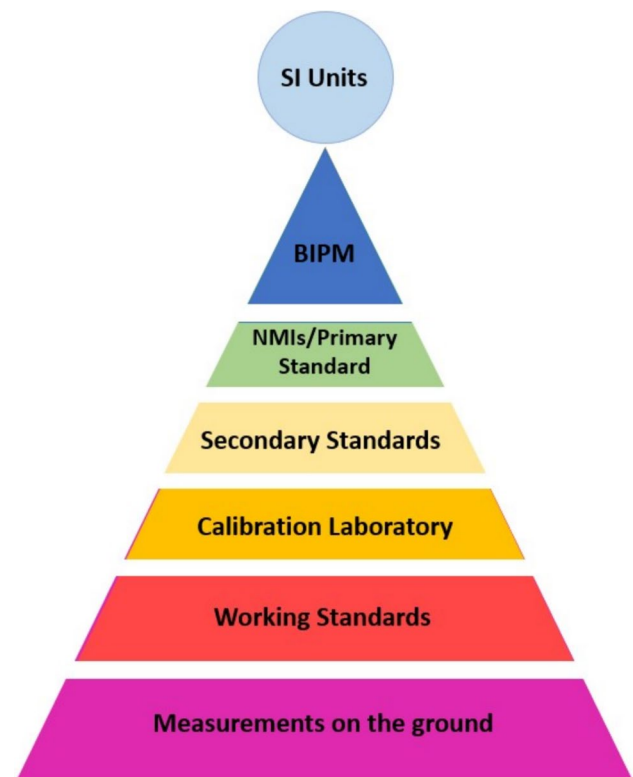


Fig. 1 The pyramid of traceability. As we move down the pyramid, the associated measurement techniques have increased uncertainty. Depending on application and quantity of interest, some intermediate steps may be missing

ISO⁸ 17025 (the international standard for testing and calibration laboratories) [40], again ensuring international consistency, this time in procedure, quality of workflow, and supporting systems.

Uncertainty quantification

Uncertainty quantification requires understanding the entire measurement pipeline and quantifying the contributions to the overall uncertainty from each step of the process. A simple example of this would be measuring the length of an item. A simple approach using a ruler would require a) knowledge of a reference standard, such as a gauge block of known size, b) understanding the bias introduced to the ruler by changes in temperature, including a measurement of the temperature and its own uncertainty, and c) the limits of precision on the marks on the ruler. Each of these contributes to uncertainty. Stated together these form what is known as an uncertainty budget [22, 23]. This also considers correlations between effects, which are important when the contributing uncertainties are combined to get the overall uncertainty.

⁷ Metrology institutions in China form a hierarchical system consisting of three levels from bottom to top, municipal, provincial, and the national institute. Provincial government can approve provincial or municipal institutes to draft calibration specifications which are used for medical devices evaluation locally as needed [38]. The national institute, also known as National Institute of Metrology (NIM), is responsible for drafting most of national calibration specifications used for evaluating medical devices' performance and traceability nationwide [39]. When national calibration specifications are issued, these supersede regional and municipal.

⁸ International Standards Organisation.

Standard approaches exist to assemble an uncertainty budget, evaluate, and combine contributing uncertainties together. In some instances (such as the one above) this can be calculated analytically, but in other, more complicated cases a better approach is to obtain uncertainties numerically via a Monte Carlo calculation, and this is the approach usually taken in practical applications [23].

Evaluating uncertainty can be challenging, and there is more than one approach to it. One option is to take a “line-by-line” approach where each source of uncertainty is identified and quantified and then combined into a single, overall uncertainty for the complete measurement process. This breakdown of contributing factors is the uncertainty budget [22, 23]. This is the approach taken historically in many areas of measurement science, and it is the approach emphasized in most metrological literature since it gives the clearest insight into the measurement process. It also identifies the sources of uncertainty which contribute the most strongly to the overall uncertainty so that these can be analyzed and ameliorated in future updates. Using this approach, it is also possible to see the overall uncertainty converge on a limiting value, which can also inform strategy for improving the measurement performance.

In some fields, however, the line-by-line approach is not practical. A good example is analytical chemistry, where measurement processes can be extremely complicated (including objects such as spectra formed in several different ways and highly non-trivial model-fitting). Additionally, sources of uncertainty can be too difficult or too numerous to characterize individually. Under such circumstances it is possible instead to work with an overall uncertainty on a final measurement rather than break it down into individual contributions. This approach has proved extremely useful and there is considerable existing literature on it, including a suite of ISO standards (ISO 5725) which may prove useful in a qMRI context.

Of course, these two approaches are not mutually exclusive. An overall uncertainty can be pursued using a combination of both methods: quantifying individual sources of uncertainty where possible and comparing their contribution to the overall measured uncertainty to give an idea of how much of the uncertainty is being captured in this analysis, which gives a measure of how much of the measurement uncertainty it is possible to quantify in detail.

In this context it is possible to identify two types of uncertainty: type-A and type-B. Type-A uncertainty is one derived from a quantitative statistical analysis of repeated measurements, whereas type-B uncertainty is one derived by other means e.g., a value from literature, a consensus value, or even a best-guess. Ideally it should be possible to break down a single type-A uncertainty for the overall measurement pipeline into individual components and obtain consistent estimates. This is not always possible, however, and

has led to the concept of “dark uncertainty”. A more extensive discussion of these concepts can be found in [41].

Metrology for radiation dosimetry

As an example of metrology in medical physics and healthcare, this section briefly sketches the role of metrology in treatment planning for radiotherapy. This is probably the most well-developed metrological framework in Medical Physics. Radiotherapy involves delivering a highly controlled dose of ionizing radiation to a specific target location in a patient’s body, maximizing the dose to the tumor while minimizing the dose to surrounding, healthy tissue. Both the amount of radiation delivered and the shape and location of the region where it is delivered are very tightly controlled and as such require very demanding uncertainties and equipment calibration [42].

NMIs are very closely involved in calibrating and maintaining performance in the linear accelerator (linac) hardware used to plan and deliver radiotherapy. Absorbed dose is traceable to the derived unit the grey (J kg^{-1}), which is the energy absorbed in a medium, manifesting as the temperature change measured by a primary standard calorimeter. Since the relevant derived unit through the gray is the joule, it ultimately necessitates traceability to the metre, second, and kilogram. NMIs typically maintain a primary standard calorimeter, which is used to calibrate secondary standard instruments such as ionization chambers which, in turn, are distributed to clinical teams to calibrate their dosimeters. Clinical teams are also audited regularly on their treatment planning procedures to establish how successful their planning and treatment delivery processes are in delivering radiation of the specified dose to the specified location and not elsewhere. This is performed using traceable, often anthropomorphic, phantoms.

This metrological framework directly underpins the delivery of radiotherapy, and therefore directly supports not just patient safety but also treatment outcome. The application of metrology gives patients and clinicians confidence that the data used to formulate treatment plans is reliable, and that the dose as delivered will not be biased from what is specified.

Quantitative MRI and why it is different

Metrology for radiation dosimetry is motivated by safety and standard of care and as such is embedded in legal responsibilities of healthcare providers. The situation in MRI is quite different. In this section we will give an overview of the motivation for improved metrology for MRI and some of the potential benefits.

In standard clinical MR imaging, image intensity data is relative, and it holds little to no information about the

numerical value of a physical parameter. qMRI goes further than this and uses the scanner (sometimes with additional tools) to make measurements. An obvious question at this point is how well the scanner is performing these qMRI measurements. In particular, what is the uncertainty associated with parameter estimates, are the estimates biased, and how similar are they to measurements made on another device? Without this information, the measurements cannot be compared, and it would be unrealistic to expect consistency between scanners or approaches. With it, however, they certainly can, even if they are made using different approaches on different hardware. Crucially, measurements are rarely (if ever) consistent as a matter of fact, and the history of metrology has shown that careful analysis and referencing is necessary before reaching the point at which measurements are consistent. MRI-based measurement is no different: qMRI needs a metrological framework that is not necessary for purely qualitative imaging approaches [43].

Any measurement of a quantitative parameter that directly relates to an indicator of normal or pathogenic function is called a biomarker. When an MRI measurand (i.e., a property being measured [22]) is used as a biomarker, it becomes even more important that its performance is understood, and its consistency maximized. Clinically, biomarkers are used to diagnose and track the progress of disease [44], and clinical trials can use biomarkers as surrogate endpoints [45]. Whilst in research there are countless MRI biomarkers in development, there are relatively few found in clinical practice or trials [46]. This was previously true of biomarkers from other fields as well—blood tests and other biochemical approaches went through decades of development and several major scandals before a metrologically robust quality control framework was developed to make them reliable, reproducible, and robust [47]. The level of quality assurance required for a biomarker to be routinely usable is high, and it takes effort and resources to develop [48]. Previous and current efforts such as the Radiological Society of North America's Quantitative Imaging Biomarker Alliance (RSNA-QIBA), the European Imaging Biomarker Alliance (EIBALL), and the Japan Quantitative Imaging Biomarker Alliance (J-QIBA) have begun work here, and already made significant progress, but there is still a need for additional metrological support.

The metrology challenges of MRI

There are three main reasons why developing effective metrology for MRI is challenging. The first is the complexity of MRI-based measurement processes. During an acquisition, an MRI scanner repeatedly plays out a pulse sequence to induce an observable signal and sensitize it to a particular physical process of interest. This can be more or less complicated depending on the contrast mechanism

and will typically also contain features such as crusher gradients designed to minimize and eliminate other, undesirable effects that would otherwise contribute to the signal. In addition, an imaging sequence is applied, where spatial relationships are encoded and contributions to the signal targeted on subregions of the image. This leads to a signal received via classical induction in a coil which describes the image in k-space. This is then transformed to image space, further processed using a variety of different techniques, and then used as input to some form of image processing to extract the actual quantities of interest [9]. Different choices can be made at all stages of this pipeline, some within the control and visibility of a typical user and some not, some revealed by a particular manufacturer and some not. This leads to hundreds or even thousands of parameters to be defined and set, but to understand MRI-based measurement, we need to understand how each of these steps of the pipeline performs. This is further complicated by the proprietary nature of much of the software provided alongside hardware installations.

The second challenge for MRI metrology is the number of different types of measurement we can make. The above pipeline is a generic description of MRI acquisition and measurement, but there are many different versions of the pipeline, most of which are not particularly standardized. We need to understand not just one highly complex measurement pipeline, but many [46]. This is highly challenging metrologically: how can we design a framework that supports so many measurands whilst allowing for the range of choices made in typical MR acquisitions and processing? As we will see, this question is at least partially answered, but further development is needed to place things on a sustainable and extendable footing.

The third and arguably the toughest challenge is that the purpose of MRI, modality and hardware alike, is clinical diagnosis. Metrology always comes at a cost, be it time, effort, or money, and at the end of the day it is always physicians and not physicists who must decide upon the cost–benefit ratio of quantitative MRI. Only where the benefit is considered high enough will the budget for MRI metrology be available.

The metrology we have

Safety and SAR

Historically, the area in which metrology has contributed the most to MRI is safety. This manifests in three ways: (1) ensuring that the energy delivered to a patient during a scan is kept within acceptable limits, (2) checking the compatibility of implanted medical devices, and (3) exposure to static magnetic fields and gradients of both patients and staff.

There has been considerable work done on these exposures at the European level. Before the publication of the European Directive 2013/35/EU on workers' exposure to electromagnetic fields [49], attention was raised on the exposure due to the movement of MRI operators through the stray stationary magnetic field of the scanners. INRIM developed dedicated computational tools to estimate the electric field induced in the moving body [50–54] and used these to calculate the safety metrics recommended in the relevant International Commission for Non-ionizing Radiation Protection (ICNIRP) Guidelines [55] for a number of realistic scenarios [56].

Additionally, MRI scanning is accompanied by safety risks for the patient that need to be controlled. First, these risks need to be quantified. The applied radiofrequency (RF) pulses lead to absorption of RF energy in tissue, the so-called specific absorption rate (SAR). SAR levels need to be limited to guarantee safe operation without excessive tissue heating. While generally well-regulated through standards [57], not all scenarios are necessarily covered and many need further investigation.

For example, parallel transmit systems use RF coils with multiple, independent transmit elements to shape the RF field for a dedicated application. Calculating the SAR per patient is challenging and is the subject of ongoing work [58–64]. Novel safety concepts such as sensor-equipped implants allow individual determination of the RF heating threat for the actual patient in the scanner could become a solution and are being investigated [25, 58–63, 65–68]. This can be performed in simulations [65, 69], via bench measurements at different frequencies [69], or within the MRI system. The effect produced by gradient and/or RF field in the body of patients bearing implants have also been extensively investigated, especially in the case of bulky orthopedic prostheses [26, 70–83].

Numerical tools for the evaluation of SAR have been developed [84, 85] including experimental validation studies [86–88] and correlation between SAR and temperature increase [89–91]. Simulations have been applied to investigate the use of metamaterials to make implants suitable to be scanned without producing artifacts [92, 93]. An alternate simulation framework was developed by NIM, which included an efficient whole-body individual modeling method with software SEMCAD from Speag [94, 95] and associated SAR evaluation as recommended in existing standards (IEC 60601–2-33:2022 [57] and NEMA⁹ MS 8–2016 [96]). A review of implant safety and related heating effects can be found in [25].

⁹ National Electrical Manufacturers Association.

The standards landscape

This section provides a general outline of how to support qMRI with documentary standards. The goal is to benefit the patient with reliable qMRI methods via a framework that demonstrates the reliability of qMRI methods at all stages. It is hoped that some qMRI methods have now reached the requisite level of developmental maturity and would benefit from a documentary standard. Beyond the legally mandated responsibilities between device vendors and regulators, the ownership for the various components discussed below may be shared across clinical sites, device vendors and regulators. Such details will inevitably evolve as qMRI moves into clinical practice.

Both documentary and reference standards symbiotically support the continued growth and development of qMRI. Documentary standards (e.g., IEC 62464–1 [97] or the IEC 60601 series [98], both of which are discussed later), as used for medical device regulatory purposes, define recognized measurement test methods and can also provide requirements or specifications for acceptance criteria. Such measurements may then be used by medical device vendors to support marketing claims to regulatory bodies (e.g., the Food and Drug Administration (FDA) in the US, notified bodies in the EU, or the National Medical Products Administration (NMPA) in China) for approval. Compliance with documentary standards requires that measurement performance is demonstrated unambiguously, and as such well-characterized and traceable reference standards are key to demonstrating that devices are performing to current standards and meet regulatory expectation. Thus, the use of both types of standards should provide the confidence that qMRI measurements enable medical staff to make the necessary care decisions.

Documentary standards typically cascade from the international level to regional and national implementations. The major international standards-setting body relevant to MRI is the IEC, which maintains a network of committees with different themes, subcommittees, and working groups staffed by subject matter experts. Members of committees are drawn internationally from practitioners, manufacturers, engineers, and regulatory bodies. International committees are typically mirrored at the national and regional levels to allow input on new standards and proposals from as broad a stakeholder group as possible. Other bodies, such as the Institute for Electrical and Electronics Engineers (IEEE) and NEMA are also recognized by regulators in some cases, but for MRI and qMRI the main body is the IEC.

The acceptance of standards at the regional and national levels varies by regulatory jurisdiction. In the EU, for example, directives and regulations like the Medical Device Regulation (MDR) EU 2017/745 define *essential requirements* a product must fulfil before it can be placed on the

market. Conformity with these requirements is automatically assumed if the device complies with applicable *harmonized* standards. The term ‘harmonized’ means the respective standard has been checked and approved at the EU level so that their structure and wording are consistent with other EU legislation, regulation, and directives [49, 99]. Other nations and regions work directly with the IEC wordings but also issue national and local directives to tune it to the specific context. This is the case in China, for example, where standards are augmented with additional Chinese-language documentation and guidance [100]. Most countries have a local standards body that mirrors relevant subcommittees on the global level. Some examples of these are the British Standards Institute (BSI), Korean Standards Association (KSA), and the Italian Electrotechnical Commission (CEI). The activities of these can vary, but typically they include local publication of international standards as well as representing national-level issues in the dissemination of international standards. Issuing a new or revised standard requires a positive vote from the IEC member countries.

Current standards

The major international standards relevant to MRI are the IEC 60601 series [57], which is a safety standard covering all medical electrical devices, and IEC 62464–1 [97] which provides specific image quality standards for MRI (but not quantitative MRI). The IEC 60601 series is a very broad collection covering a wide range of devices (in principle, any electrical device used in a healthcare setting) and covers safety issues with careful attention paid to the scope of individual devices and which parts of the standard are relevant to them. The series consists of a base standard IEC 60601–1, applying to all devices, 12 *collaterals* (IEC 60601–1-X) about certain specific, but not device specific, safety and performance aspects, and a list of around 80 *particulars* IEC 60601–2-Y, dealing with one specific device type. The particular standard for MRI is IEC 60601–2-33 “*Particular requirements for the basic safety and essential performance of magnetic resonance equipment for medical diagnosis*”. It is of such crucial importance, that it is often simply referred to as *the* IEC standard in the MRI community.

The IEC 60601 series is a broader standard than just MRI and applies to other medical devices as well as MRI. Some sections, such as IEC 60601–2-33 pertain to MRI only [98] and enhance or supersede the general standard. As currently defined, the MRI standards does not specifically cover qMRI approaches, although there is interest in developing coverage where appropriate and once techniques are of sufficient maturity to warrant it. It is also worth noting that the user manual for a typical MRI scanner often states that the scanner is not a measurement device [101], a fact that is often surprising to researchers developing quantitative methods,

who are known to feel differently [102]. “Special diagnostic procedures” using MRI are also not addressed in the scope of IEC 62464–1. Standards are also under constant review and undergo periodic redrafting exercises. At the time of writing, IEC 60601–1 is undergoing a substantial revision [51], and revisions to IEC 62464–1 are being contemplated.

A further series of standards from American Society for Testing and Materials (ASTM) International deals with the safety aspects in the context of foreign objects in or near the MR scanner. Such items can be external, e.g. a physiological monitoring device with electrodes attached to the patient’s skin, or internal, like a medical implant. ASTM F2503 was groundbreaking in introducing the terminology which was later adopted in many other standards. Foreign objects are classified as i) *MR unsafe* (“an item that is known to pose hazards in all MRI environments”), *MR conditional* (“an item that has been demonstrated to pose no known hazards in a specified MRI environment with specified conditions of use”), or *MR safe* (“an item that poses no known hazards in all MRI environments”). Importantly, image artifacts are considered (diagnostic) safety hazards which automatically excludes all metallic devices from an *MR safe* label. ASTM F2052 and F2213 define test methods to investigate the magnetic displacement force and torque, respectively, on medical devices in an MRI environment. Such investigations are highly relevant, safety-wise, but technically more or less straightforward and not an active field of metrological research. ASTM F2182, entitled *Measurement of Radio Frequency Induced Heating On or Near Passive Implants During Magnetic Resonance Imaging*, describes standardized test methods and equipment to investigate the thermal response of passive metallic implants to RF exposure from the MRI scanner. These ASTM standards were co-developed by the US FDA and are tailored to provide the information FDA needs during the medical device review process. The procedures described in ASTM F2182 for passive implants are also helpful to perform validation experiments as required by the ISO/TS 10974 standard for assessing the hazard from active implants in an MRI environment, issued by the International Organization for Standardization (ISO).

ISO/TS 10974 provides manufacturers of active implants with defined procedures and methods to assess the MRI compatibility of their devices. In many respects it is the active-implants counterpart of the ASTM series on passive implants. The standard, for some tests, applies a tiered approach with simple assessment procedures but large safety margins in Tier 1, allowing a simple and cheap investigation of implants with little hazard potential, up to the extensive (and expensive) full-scale assessment of Tier 4 as in the case of RF-related evaluations.

Medical device vendors both “verify” and “validate” the device operation. Verification is an engineering concept that confirms the device under test (DUT) meets specified

requirements. Validation is a demonstration that the DUT fulfils the requirements of its intended use (e.g., the confirmation that a medical device meets a clinical need). For example, a test object purporting to replicate the diffusion value range of a particular organ has been verified and validated. If such a connection between test object and clinical need does not exist, there is little or no proof that the clinical measurement is reliable in the absence of the test object. Collecting the clinical and test object measurement simultaneously is a possible solution. If the test object measurements do not meet expectation, then the clinical measurements may be suspect.

In addition to the international and national documentary standards, there are also recommendations produced by various professional bodies, such as the American College of Radiology (ACR) and the American Association of Physicists in Medicine (AAPM). Whilst the national- and international-level standards are often entwined in regulatory and legal metrology, it is worth noting that recommendations produced by these other bodies do not confer any legal obligation or responsibility regarding the standard of care recommended, see [103]. The ACR [104] provide for both a QA service and also a guidance document on procedures for the testing, which shall be discussed later. American Association of Physicists in Medicine (AAPM) report 100 [105] was developed as an aid to acceptance testing, and offers guidance on a variety of MR system checks as well as mechanical ones, along with indicative performance thresholds/acceptance criteria. The Institute of Physics and Engineering in Medicine (IPEM) produced an extensive report on a wide range of performance tests in Report 112; “Quality Control and Artefacts in Magnetic Resonance Imaging” [104, 106]. Both this document and the suite of relevant MRI Standards from NEMA provide comprehensive detail on a variety of objective tests suitable for quality control and accreditation. It is worth noting, however, that the field of qMRI is not especially well represented in many of the guidance documents available, likely a consequence of the nascency of the field of metrology in these areas.

Potential future development

For qMRI purposes, a documentary standard could take various forms. One option is that it could specify a defined test method, including test objects. This would originate from a committee of experts in the relevant fields finding broad consensus on a methodology representative of the best possible realization of the measurand in question and defining the appropriate route to traceability. Alternatively, a less restrictive but no less impactful standard could take the form of defining the minimum level of fidelity to which a measurement must be made. An example of this may be seen in the context of radiotherapy delivery where the

International Atomic Energy Agency (IAEA) published a recommendation on the level of uncertainty that should be reached for dose delivery [107].

It is essential that the qMRI test method measurement provides a clinically relevant value. A test object and method that provide a measurement which meets the output expectation of all experts is useless unless it sufficiently mimics clinical reality. Without this it cannot provide reasonable confidence that patient measurements are, by extension, reliable. If a test object and typical human volunteer travelled to many different MR scanners, it is a requirement that the qMRI measurements from both be reliably correct, and the test object measurement must be a reasonable facsimile of human tissue.

Note that while the documentary standard test methods need to be well-defined, and some methodology implementation variance is tolerated, test objects are typically allowed a large range of design options. This requires a careful itemization of all the test object requirements in the documentary standard, but potentially allows for different implementations (e.g., a small versus large version of the phantom in consideration of the range of receive coil sizes). Documentary standards should not be used to promote business interests by mandating specific test methods and test objects, which may be patented. If there is a patent position on some element of a documentary standard, it must be identified via a declaration in the standard.

Artificial intelligence (AI) methods and applications are rapidly evolving, and the implications for qMRI must be considered. Given a regularly imaged test object, the AI-derived qMRI measures may vary over time and across software updates. It is assumed here that a typical constancy test comprising of traditional image quality (IQ) metrics might not detect changes in qMRI measures where AI methods are involved. While a qMRI constancy test needs to identify such temporal variations, the clinical implications of the variations will be unique to the clinical need. What constitutes an acceptable level of qMRI variation may be a site-specific determination. In this regard, a qMRI constancy test has the benefit of an expected target value.

The basic premise that qMRI outputs will influence clinical decision making and perhaps reduce the scope or reduce the need for image interpretation means that if the numeric output is incorrect, inappropriate clinical decisions are possible, and “essential performance” may be at risk. It may be important that the qMRI outputs are demonstrated to be correct in a fashion similar to proving that a medical device emitting radiation is delivering the intended dose to the intended location. Quality assurance and control (QA/QC) of radiation delivery is an excellent example of a well-defined methodology producing measurements that are

traceable to reference standards and presumably confirming the outputs are as intended.

It is also essential that the MR scanner itself be maintained to its expected specifications. Traditional IQ metrics-based constancy tests are a helpful tool. A recently restored quenched magnet may no longer have the identical B0 shim due to subtle shifts in the magnet windings, or an eddy current signature may change because the magnet internal temperature shields are now at a slightly different temperature after a helium refill. A gradient amplifier swap may require a pre-emphasis recalibration. All parameters may be within specification but there could still be operational differences in the context of historical continuity. Such constancy tests are ideally quick and simple.

Mature standards are necessary to demonstrate the compliance of a technique to the best available knowledge of how to perform it. There will always be a need for this capacity, and as the size of the field grows so too does the need for appropriate evolution of international standards. Quantitative MRI has the potential to proliferate significantly into clinical usage, and it is vital to ensure the standards landscape enables the translation of promising diagnostic and therapeutic methods.

Phantoms, new and old

Phantoms play an important role in evaluating MRI system performance, historically providing a controlled method to assess image quality such as contrast resolution, spatial resolution, signal-to-noise ratio (SNR), and geometric distortion. Phantoms are test objects which can be used to evaluate a particular aspect of the image formation or measurement process by providing a suitable reference.

One of the most commonly used phantoms is the ACR MRI accreditation phantom, which was introduced in 1992 and evaluates geometry distortion, spatial resolution, and slice thickness [108, 109] [104, 110]. The ACR phantom was primarily designed for head imaging coils, but it presents significant limitations when applied to other RF coil types, such as breast and knee imaging coils. The Magphan phantom introduced in 2010 provides enhanced capabilities for evaluating geometric distortion, particularly in neuroimaging [111–113]. However, these phantoms are also constrained by their fixed configurations. It is worth noting that these phantoms are used for evaluation of basic system performance in preparation for traditional qualitative investigations. Many of the tests themselves require subjective procedures [104], where different operators would produce differing results based on their levels of experience, access to support software, etc. Any quantitative measurement of spatial characteristics (e.g., distortion, volume, etc.) would need to have appropriate metrology to characterize any

inherent variability in the reproducibility of the testing, as well as traceability in the phantom.

In practice, the level of sophistication of the design and construction of a phantom depends on what it is used for. Phantoms to assess image uniformity by eye may be as simple as bottles of water [114]. Conversely, a phantom used to benchmark the performance of a quantitative imaging process should ideally have traceably verified reference values with accompanying statements of uncertainty [4]. In metrological terms, a traceable phantom is a form of reference standard, providing one or more reference values to assess a scanner's measurement capability. Fully metrological usage of phantoms in qMRI is comparatively recent, though it is becoming more widespread, and there is also an intermediate class of phantoms which aim to provide reliable reference values without providing full traceability.

System phantoms [115, 116] are multi-purpose quantitative phantoms that contain test media (usually gels or solutions doped with paramagnetic ions) to represent a variety of target measurands such as T1, T2, and ADC. Other phantoms are made bespoke to their intended purpose, often homemade, to support local research efforts, or specifically designed for a certain body region, e.g. Alzheimer's Disease Neuroimaging Initiative (ADNI) phantom for the brain [111] or the T1-mapping and extracellular volume standardisation (TIMES) phantom for cardiac imaging [117–119]. Unfortunately, there is little capability to independently benchmark the properties of these phantoms, nor is it possible to separate out scanner variation from unexpected imperfections in the behavior of the object.

The increasing complexity of MRI systems, driven by advancements such as wireless, flexible, and adaptive RF coils, has created a need for more versatile phantoms. This has led to the development of modular designs, such as the LEGO-compatible Modular Mapping (MOMA) phantom [120] or the iMet-MRI Phantom [116, 121] offering customizable configurations suited to different anatomical regions and coil types [120]. Modular phantoms present a flexible solution for comprehensive MRI quality assurance and quantitative imaging assessments, effectively overcoming the limitations of traditional fixed phantoms. There is also an emerging class of “virtual phantoms” (e.g., [122]) which work by injecting an RF signal into the receiver coil during a scan.

There are various commercial companies who specialize in the manufacturing of phantoms for quantitative MRI¹⁰ and some of these put their products through rigorous metrological evaluation to ensure the high quality and stability of their test objects. This allows for the expected behavior

¹⁰ Including but not limited to: Diagnostics, Poland. Gold Standard Phantoms, UK. CaliberMRI, US. Calimetrix, US. Phantom Laboratory, US. Sun Nuclear, US.

of these phantoms to be part of the traceability chain and guarantee a level of confidence to the end user.

In addition to physical phantoms, digital reference objects and numerical simulations can be a valuable tool. Digital reference objects are mathematical models which capture the details of the measurement process as well as a model of the underlying physics of signal formation [123]. By considering the measurement process and underlying physics *in silico*, each aspect of the process can be controlled and varied in a way that is not always possible using physical objects. This is valuable in understanding image quality and also in evaluating uncertainties.

A related concept is the digital twin [124], which is a detailed simulation of a device, for example an entire MRI scanner, which aims to be a digital copy of a specific device and is updated as the device is updated to maintain correspondence. Digital twins can also be used to understand measurement performance and evaluate uncertainty and are particularly useful for identifying where important effects contributing to uncertainty or bias may have been overlooked by comparing the digital twin's performance with that of the live device.

Electrical properties of tissue

Electric properties tomography (EPT) is a family of qMRI techniques that aim to obtain an indirect measurement of tissue dielectric properties (such as electrical conductivity and permittivity). Interest in these quantities is motivated by the possibility to exploit them as biomarkers (in many types of cancer, the values of conductivity and permittivity tend to increase with respect to the corresponding healthy tissues). Information about these parameters can be used to optimize therapies based on the application of electromagnetic fields and to perform subject-specific SAR assessments.

Work in this area has been led by INRIM, and includes a comparison of EPT techniques [125], uncertainty analysis [126–131], and development of tissue-mimicking materials [132]. A software library is available for the approach (EPTlib—<https://eptlib.github.io/>) [133], and a study of the tissue dielectric properties, water content, and T1 [134]. The technique's repeatability and reproducibility have also been investigated using a saline phantom [135].

Proton density fat fraction

Proton density fat fraction (PDFF) is a widespread technique in the research literature and is also gaining popularity in clinical applications [136, 137]. The technique aims to objectively measure the overall amount of fat in tissue, a parameter which provides information about several pathologies. PDFF is also amenable to phantoms, since the overall amount of fat can be controlled during the preparation of

samples. Whilst several commercial fat fraction phantoms exist, many with highly sophisticated materials designed to mimic the NMR spectra of human fat, as far as the authors are aware, none are SI traceable.

To diagnose fat-related diseases such as metabolic dysfunction-associated fatty liver disease (MAFLD) [138] it is essential to quantitatively measure fat content in tissue. PDFF is considered a non-invasive, quantitative imaging biomarker for fat content estimation and is being utilized not only for diagnosis but also as a primary outcome in clinical trials for MAFLD-related drug development [139]. Due to differences in implementation across vendors, ensuring measurement repeatability and reproducibility is critical when acquiring multi-center data. To achieve this, water-fat emulsion phantoms are often created [140, 141], or commercial phantoms [142–144] are used. However, these phantoms are not SI traceable and are manufactured with various additives, leading to a lack of data on the chemical fat fraction, homogeneity, and stability of the materials. Therefore, there is a need for reference materials/phantoms with SI traceable homogeneity and stability.

The only NMI currently offering surfactant-free water-fat emulsions as reference materials for fat fraction imaging is KRISS, who have developed approaches for PDFF as part of their modular phantom approach [145], and chemical metrology for fat fraction imaging using a simple fat–water solution was also developed in the iMet-MRI project [116, 121]. In both cases samples with controlled fat and water content are prepared with calibrated balances and using standard reference materials traceable to the Kilogram.

Key quantitative MRI capabilities

This section summarizes key metrological capabilities either available directly from national measurement institutes or aimed at underpinning services in the future.

KRISS reference materials service

KRISS have developed reference materials for qMRI including SI-traceable solutions suitable for T1 and T2 mapping and a surfactant-free emulsified reference material to improve MRI-PDFF (proton density fat fraction) accuracy for multi-site and multi-vendor evaluations [146]. The materials are contained in a modular, Lego-compatible phantom which can be assembled into many different configurations depending on the application and choice of receiver coil. Full details of the approach can be found in [145], which also characterizes homogeneity and stability of the materials. KRISS offer access to these modular phantoms and guidance on their use as a commercial measurement service

and they are currently the only NMI to offer reference materials for PDFF approaches.

The NIST MRI biomarker measurement service

The NIST MRI biomarker service offers traceable measurements of T1, T2, and ADC of material samples at specified field strengths (1.5 T or 3 T) and temperatures. These are provided with fully characterized uncertainties. At present, the measurement service provides calibration of proton relaxation times [147] and isotropic diffusion coefficients [148]. Other biomarkers, where traceability has not yet been established, can be measured as part of a special test where the full measurement protocol is documented and an error is estimated, but a well-defined uncertainty is not provided. For an extensive discussion on the details of the measurement service and traceability, see Ref. [149]. The NIST MRI biomarker measurement service is based on a dedicated metrology NMR system shown in Fig. 2. The system can vary the sample temperature over the range of 0 °C to 60 °C with a 0.2 °C uncertainty. It has 3-axis gradients, a variable magnetic field from 0.5 T to 7.0 T, and a research console (Tecmag Redstone) that allows easy generation and verification of both nuclear magnetic resonance (NMR) and MRI pulse sequences.

For MRI measurements, the chain of calibrations extends back to the definitions of time, length, and temperature. One example of a traceability path is for MRI diffusion measurements, which require a precise measurement of magnetic gradient strength. This is done by taking a MR image of an object with calibrated dimensions, measuring the width in the frequency domain, and then calculating the gradient strength. Here the traceability occurs through calibrating the NMR time base against the NIST cesium atomic fountain clocks and measuring the object dimension using a micrometer calibrated with traceable gauge blocks. Uncertainty is determined using a Monte Carlo Bloch Simulator [149].

Phantom libraries

To improve access to phantoms and maximize their usage in the field, both NIST, in partnership with the US National Institute of Biomedical Imaging and Bioengineering, and NPL have phantom lending libraries. These libraries contain commercially available phantoms for multiple MRI measurements including relaxivities, ADC, chemical composition, and more specialist test objects such as diffusion microstructure and kurtosis phantoms. Phantom libraries typically lend items for a fixed period at low or minimal cost, and offer

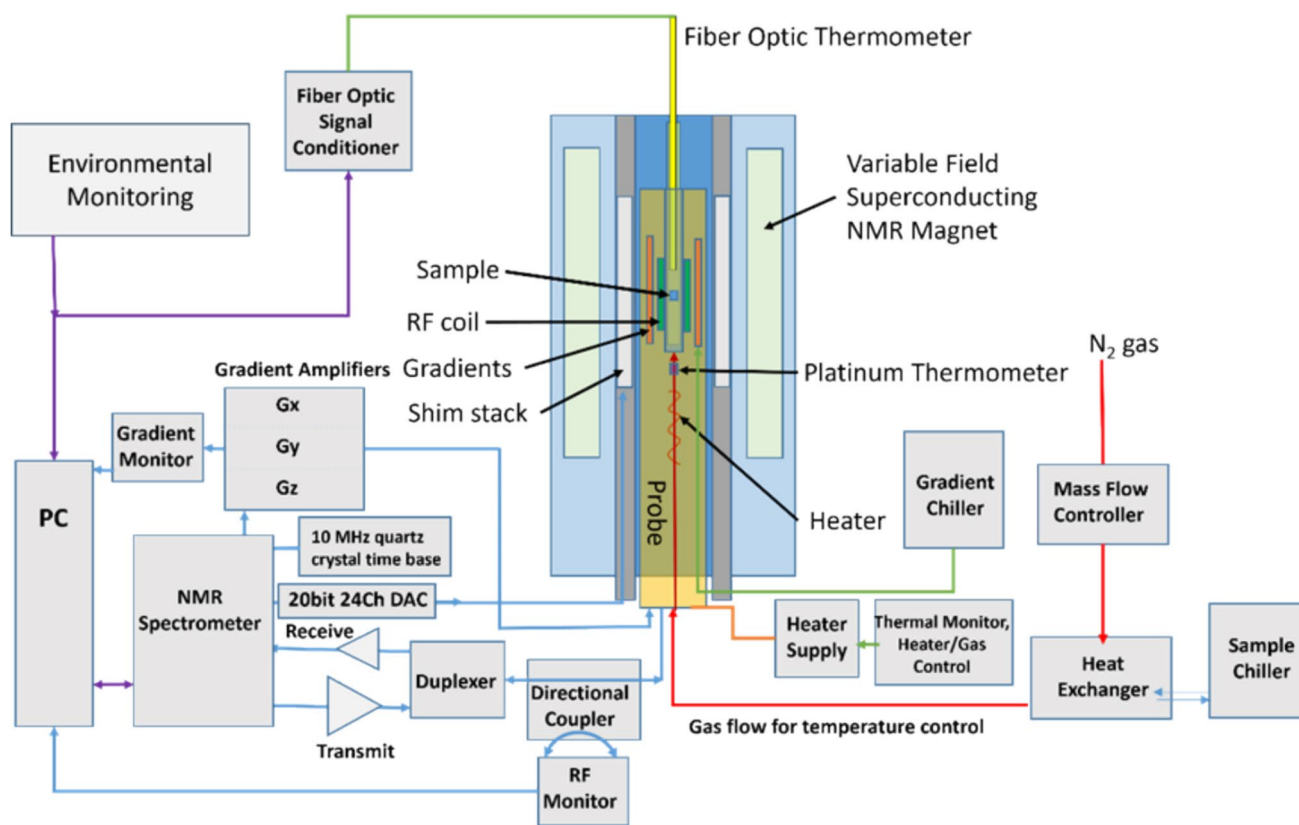


Fig. 2 Schematic of the NMR system used for NIST MRI Biomarker Measurement Service

advice on how to use, acquire, and process phantom image data on typical MRI systems.

The iMet-MRI project

The development of fully traceable MRI metrology in Europe was largely through the iMet-MRI project. This was a coalition of five European measurement institutes (INRIM, Türkiye Bilimsel ve Teknolojik Kurumu (TUBITAK), the Institute of Metrology of Bosnia and Herzegovina (IMBiH), and the UK National Measurement Laboratory hosted at the Laboratory of the Government Chemist (NMI at LGC), and led by NPL) which developed traceable materials characterization methods for T1, T2, T2*, ADC, and PDFF including associated chemical metrology for sample preparation and analysis. The project also developed image acquisition protocol and analysis software, demonstrated their performance in an international multi-site trial which included repeated measurements on 1.5 T and 3 T scanners from Siemens, Philips, and GE, and made acquisitions at both 7 T and 0.55 T. Several sets of materials were synthesized, and the results are in the process of being published [116, 121]. iMet-MRI capability is now in the process of being turned into a measurement service facility at NPL using a variable field spectrometer similar to the one used at NIST. An associated metrological intercomparison of the two capabilities is planned once the new facility is complete.

The metrology we need

Thus far we have concentrated on the current developments of MRI metrology. We have shown that there is an existing body of knowledge and set of expertise globally that provides and underpins measurement capability for MRI safety, and there is also the capability for SI traceability in a small set of MRI measurands via a single primary standard facility at NIST. In this section, we take a more prospective viewpoint and compare current MRI metrology to a fully mature metrological framework, considering the potential benefits and outlining future activities that could take us there.

Challenges with the current framework

A key issue with current MRI metrology is capacity. The currently employed model for traceability involves commercially manufactured qMRI phantoms containing materials characterized directly on a single primary standard located at NIST. This means that to manufacture traceable phantoms, samples must be sent to NIST for characterization. Since there is an unavoidable variation in batches of materials, this must typically be done on a batch-by-batch basis. In metrological terminology, qMRI phantoms are

transfer standards that allow scanner performance to be assessed, but (notably) not calibrate its output.

This arrangement as it stands has the advantage of providing measurements via the smallest possible number of steps between primary standard and measurement-on-the-ground, but suffers from problems of expense, capacity, and fragility. Specifically, primary standard measurement services are expensive, and the necessity of performing batch-by-batch characterization leads to substantial costs for phantom manufacturers which are, of necessity, passed on to customers. This means traceable MRI phantoms are expensive, with little scope for cost reductions through scale or efficiencies in other parts of the manufacturing process. This frequently places phantoms beyond the financial reach of typical clinical imaging departments, even in countries with well-developed healthcare systems, and makes them very much beyond the reach of imaging facilities in less developed nations. This is a significant barrier to the translation and uptake of quantitative MRI approaches and the associated patient benefit.

Having traceability available via a single primary standard globally also leads to risks. First, a single facility, however, efficiently run, means there is only so much capacity for providing measurement services and leads to lengthy waiting times when demand is high. Effectively, it is a cap on growth for the traceable phantom industry. A single facility also means that any technical problems at the facility result in an interruption in availability of services, potentially further lengthening waiting times. This also impacts scheduling of measurement services. If there is a lengthy downtime associated with, for example, a fault or upgrade this would increase the difficulty of scheduling measurements whilst maximizing capacity. All of these considerations exist for any key facility, but the current lack of redundancy in the system makes this risk especially acute.

Capacity issues become even more important if we also consider the need for improved international standards from qMRI. For a new standard to be effective and agreed upon, it must be possible to demonstrate compliance. Since a new standard would (at minimum) specify that measurement performance must be stated and quantitatively benchmarked, traceable measurements would be needed in far greater numbers than are currently available. The same is true for effective translation of quantitative methods into the clinic. Routine QA would be needed to monitor performance, and therefore further increase demand for objects and procedures underpinned by traceable metrology.

Finally, if there is only a single primary standard globally, it is not possible to perform comparisons of performance with other facilities. This means that the ultimate check and balance for measurement performance is not possible, and any drift in measurement performance at the primary standard is difficult to detect.

There is currently considerable support for increased capacity in traceable metrology. NIM and KRISS have developed national-level capability to improve access to traceable references and have expressed interest in increased international co-operation. NPL and INRIM are currently building their own primary standard facilities, aiming to provide similar measurement services to NIST. All of this, while encouraging, will not provide the step-change in metrological framework the field currently requires, however, and there is instead a need for a new traceability structure.

Among regional metrological organizations, Focus Group of Medical Metrology (MMFG) in the Asia–Pacific Metrology Programme (APMP) has expressed willingness to further increase members' attention to the evaluation and traceability of large medical imaging equipment, including MRI systems. Due to significant differences in economy and healthcare system among APMP members, collaboration with other international or regional organizations has been strongly suggested and welcomed to help improve construction of traceability systems and metrological capabilities among APMP members.

Secondary standards

Drawing on the experience of other fields of metrology, one way to achieve increased capacity is to develop secondary standards [28]. As discussed previously, a secondary standard is one that is calibrated directly to the primary, and there has been the suggestion that this could be a good model for MRI metrology [150]. By calibrating a set of secondary standards to existing primary standards, measurement service capacity can be rapidly increased, and costs driven down. Secondary standards for MRI do not currently exist but could be developed with appropriate international collaboration between NMIs and a set of MRI stakeholders (including hospitals, universities, calibration labs, and manufacturers). Furthermore, there is no restriction to a single configuration for secondary standards. Different types of standards suitable for different stakeholders could also be developed.

Secondary standards are typically lower cost than primary facilities, and these could be sited at interested NMIs, existing commercial calibration labs, or at manufacturer facilities themselves. This leverages existing measurement capacity to build the next layer, supporting broader uptake of qMRI methods. In other fields one of the prime advantages to using primary standard calibration, as opposed to entering the chain of traceability further down, is minimization of uncertainties. It may be that this is not as significant an issue for some applications in qMRI, as the measurement uncertainty introduced may be of an order of magnitude lower thanks to inherent patient variability.

Accessibility

A very significant challenge for current MRI metrology is access to traceable phantoms by imaging departments and facilities at local hospitals and those in countries with less well-funded healthcare systems. Traceable phantoms are both expensive, and not explicitly required by current standards and guidance [46]. To mitigate this, both NIST and NPL maintain lending libraries whereby phantoms can be borrowed by clinical and research teams. KRISS operate a similar model whereby traceable modular phantoms can be sent to imaging teams to assess scanner performance in quantitative applications, and the phantoms are returned when they are finished [151].

A phantom package in the lending library may consist of far more than just a physical phantom. Ideally the package contains an instruction manual, reference data, recommended imaging and analysis protocols, stability analysis protocols, safety data, recalibration protocols, and a digital twin. The library validates the properties of the phantom, shares data, and curates phantoms. Curation involves monitoring the stability of the phantom properties, fixing or upgrading components as needed, providing verification of both the primary and secondary properties, and maintaining availability for long-term studies.

With recent developments in low field, portable scanning technology such as the Hyperfine Swoop [152] as well as funding from the Gates Foundation, there are efforts to introduce MRI measurement capability into lower and middle income countries. This enables the delivery of improved healthcare in resource constrained areas, such as parts of Malawi [153] where there may not be the requisite infrastructure to support a full scale 1.5 T or 3 T MRI facility. Correspondingly there would not be resources available for high-cost metrology to support any qMRI measurement techniques in these settings, and developing low-cost phantoms and more access to the traceability chain only further spreads the accessibility of potential quantitative techniques.

New phantom materials

A very common criticism of current qMRI phantoms is that they are too simple and do not approximate the MR-measurable properties of tissues closely enough. Paradoxically, phantom materials can often lead to measured data which are too close a fit to the assumptions used to analyze them and as a result fail to capture important confounds which are present when applying the same approaches in vivo. A good example of this comes from diffusion imaging, where diffusion in a fluid leads to monoexponential behavior in the measured diffusion attenuation. This is exactly what is assumed by many analysis models (e.g., in calculating an ADC or fitting a diffusion tensor) but is not what is observed

in vivo. Here, the interaction between diffusing spins and the tissue environment leads to a slower than exponential signal decay and as such a dependence on measured diffusivities on the overall diffusion sensitization (or b -value) [154].

In this case and others like it, the mismatch between the modelling assumptions necessary to obtain parameter values and the actual behavior of the MR signal in vivo is a source of measurement bias, and in many applications we currently lack phantoms sophisticated-enough to evaluate it. The phantoms that do exist are not SI-traceable and many are bespoke items built for a specific research purpose, rather than widely-available products. There is a clear need for metrologically robust SI traceable phantoms which allow these types of effects to be fully characterized.

AI in quantitative MRI

In recent years AI has also become the state-of-the-art for MR image reconstruction with methods ranging from image enhancement as a post-processing step to complete image reconstruction using large fully connected networks [130, 155]. Currently the most used techniques combine AI with information about the MR acquisition in a cascaded network architecture [131]. AI approaches have also been presented for qMRI [156–159].

Like any other inference method, it is crucial to understand the uncertainties associated with AI approaches. This is more challenging in some approaches than others, and it is connected to the broader field of AI explainability and trust [160]. In any AI application in medicine, it is critical to understand how uncertain an algorithm is regarding a particular conclusion or value. AI is used in MRI in a number of different ways, from accelerated image acquisition and reconstruction to artefact removal and diagnostic inference [161]. In each case errors made by an algorithm can impact patient outcomes. It is, therefore, critical to incorporate a metrological understanding of algorithmic performance into new frameworks to support the use and deployment of AI.

Improved QA procedures

Through the use of available SI traceable phantoms, it is now possible to quantitatively assess measurement performance of individual systems, in particular qMRI applications in T1, T2, and ADC, at least in the phantoms themselves. This immediately raises the question of quality assurance. How should we use this capability to meaningfully assess measurements made in individual studies? There is currently no consensus on this, and although it is not clear currently what form such QA should take and how often it should be performed, it is a topic on which a consensus could be reached. This would require input from a broad range of stakeholders: clinical teams, technologists, clinical scientists, research

MR physicists and image analysts, as well as scanner and phantom manufacturers, standards developers, and NMIs.

A closely related question is whether and to what degree it is possible to calibrate measurements acquired in different ways on different systems. In other forms of measurement, it is common to introduce calibration factors to minimize quantified systematic bias [22, 23]. This can have a substantial effect on reproducibility. The extent to which this is possible, feasible, and tractable is an open question in qMRI and will certainly depend on the measurand of interest, but the current metrological framework does make this possible.

It is worth emphasizing that QA is not a one-size-fits-all problem. The level of QA necessary for clinical business as usual will necessarily be quite different than that used in trial qualification and ongoing monitoring, and it will be different again from what an individual research study might require. It is certain, however, that some form of QA for qMRI is necessary. Without QA we cannot meaningfully and quantitatively compare measurements made in different studies and there may be substantial gains from introducing it.

Accreditation and trial qualification

Once measurement performance can be fully quantified and benchmarked, it opens new possibilities for scanner assessment and study design. The most immediate possibility is a rigorous quantification of measurement variability for a given device in a particular application, which provides more detailed data about variability per scanner in a trial [112, 162]. The variability in measurements for a particular trial can be more readily evaluated, and as a result there is more data from which to estimate a trial statistical power than would otherwise be available. This cascade enables more confident estimates of the required number of participants for a given effect size. These estimates are currently very challenging for MRI-based assessment and leads to relatively conservative estimates of study size [163]. Quantifying variability in a scanner's measurement of a particular quantity also means that performance can be monitored and maintained between studies. Additionally, study qualification can be stated in terms of an existing level of performance, reducing the need for lengthy and repeated harmonization phases in larger studies.

Beyond measurement variability for trial qualification, recent work has pointed out that benchmarked scanner performance potentially leads to statistical criteria for measurement performance. The statistical concept of the perfect machine can be brought into an MRI context [150, 164]. The perfect machine is a concept which states that once measurement variability is reduced to a level where it is small compared with the variability of the measurands across a patient cohort, and there is no need to reduce it any further. It should be noted, however, that even when the

measurement uncertainty falls below that of natural patient variation, it still may allow for valuable information on long term scanner variation or allow for the detection of emergent faults before they escalate to a point of clinical impact.

When imaging pathology, in many cases the variability of a particular measurand across a patient population is known from non-imaging-based methods and as such enables targets to be set for measurement variability. Hall et al. [150] discuss a grading system for scanner performance in different contexts, and its feasibility goes further and postulates an accreditation scheme for scanners that could be used to recruit sites for clinical trials.

Whilst it remains to be seen whether there is appetite for such a scheme, or how easy it would be to introduce one, this is illustrative of the new possibilities associated with fully benchmarked and traceable measurement performance in MRI.

Discussion and conclusions

Metrology is a silent partner in many fields of science and engineering. NMIs function as the link between the thousands of measurements that are made every day in modern industrialized countries and a single, unified system of units that allows them to interact, trade, and compare data and results. We have seen that while metrology for MRI is still a work in progress, the principles of traceability and uncertainty quantification are in place for a small number of measurands, and the same principles are in use to ensure device compatibility and patient safety. We have also seen a strong link to the world of documentary standards, where current standards exist for basic image quality, with work underway to develop further unifying standards for qMRI.

There are currently six NMIs globally with an interest in developing MRI metrology, and there is considerable interest from all of them to work together to develop a truly international metrological framework for MRI. In some ways, MRI metrology is at a very exciting stage. Recent work on MRI standardization such as RSNA-QIBA (https://qibawiki.rsna.org/index.php/Main_Page) has highlighted a broad appetite for standardization in many areas of MRI, as have developments in vendor-neutral imaging implementations [165–167]. Improved metrology for MRI has an important role to play in underpinning these and other similar efforts by providing the link to SI units.

The challenge for MRI metrology over the coming years is to transition from the current proof-of-principle phase to a more mature and robust framework, suitable for deployment across many different countries and regions, each with their own priorities and needs. The need for improved standards has been highlighted as critical to unlocking the full potential of qMRI in clinical practice, clinical trials, and to

empower AI approaches to imaging and imaging science. To do this, it is imperative that the underpinning metrology has the capacity to support compliance with a new standard. Secondary standards are certainly part of the answer here but so are streamlined approaches to supporting new measurands and studies generating guidance for how to maximize reproducibility.

There are, of course, some questions which arise from metrological approaches to qMRI which cannot be answered by metrologists. One is the question of how equivalent different measurands actually are. For example, are the T1 relaxation times measured by different types of acquisition or fitting model genuinely the same quantity? Questions like these require consensus from the entire community, but it is possible to at least make a suggestion as a point to stimulate debate. In this example we could argue that a T1 as measured by an Inversion Recovery (IR) sequence could be used as a definition of the quantity of interest. We do not always use an IR sequence because it is very time consuming and not always practical and so alternative acquisitions have been developed that attempt to measure the same quantity in a different way, which introduces additional uncertainty and bias. From this viewpoint, alternative acquisition schemes and modelling approaches are approximations of an IR sequence and thus need separate characterization and analysis but nonetheless attempt to measure the same parameter.

An analogy might be drawn here with simpler, more well-established measurement procedures. For example, measuring electrical resistance in a wire or other component. To do this, a current must be applied, but the measured value will be affected by the magnitude of the current. A small current will cause less heating than a large one and lead to a different estimation of resistance, but the actual underlying parameter being measured is the same. The system is perturbed differently by the two different approaches and this needs to be understood, but the measurements are related and comparable. This is discussed in the VIM, Sects. 2.10 and 2.11 [22]. Of course, this is just a starting position for debate and readers may disagree (perhaps strongly).

Characterizing the performance of a given choice of acquisition, modelling strategy, etc. ideally should be part of an individual quantitative study, and in some cases, this may be challenging. Performance in phantoms is certainly part of this assessment, as is potentially the use of digital reference objects to test effects which cannot be captured in physical objects. Again, in some cases this also may be very challenging, but in many other fields this is achieved by calibration, where a device in a particular application is rigorously tested and its performance characterized. The information from a single calibration may then be used in many subsequent studies. The complexity of MRI presents a challenge here: it is certainly not possible to calibrate a scanner in and of itself, but it may be possible to perform a calibration in a

particular application. The question of to what extent this is possible is, again, one for the entire community. Metrology can only offer a single viewpoint. Asking what amount of the variability observed in multi-site in vivo studies is due to genuine uncertainty and which is due to differing biases could potentially offer a way to reduce it.

QMRI metrology is an effort that touches a large and diverse group of stakeholders and making sure all voices are heard is imperative. Clinicians, physicists, engineers, technologists/radiographers, scanner manufacturers, phantom manufacturers, regulators, clinical research organizations, pharmaceutical firms, metrology institutes, and patient groups all have their own viewpoints that need to be heard and understood as we go forward, but the same is true of previous metrology and previous standardization projects which demonstrated that this is possible.

There is also the question of who should perform the necessary QA and QC to support qMRI, once it is developed and deployed. This, again, is a topic for debate but we can look to other fields for examples. In radiotherapy, for example, QA and QC is carried out both by manufacturers demonstrating essential performance and safety and by clinical scientists and technologists to confirm and monitor performance as the systems is used. In some cases, QA tests can be automated, but this requires the automated tests to maintain calibration. In current clinical MRI, QA is also performed both by manufacturers, sometimes as part of a service contract, and by clinical science teams, and it seems that this is a basis to proceed from. There is certainly a question regarding how much and how frequent QA for qMRI should be. In research it is likely that many studies will require additional QA over and above any typical clinical routine simply because research studies are more varied than clinical day-to-day scanning. In this case it is the researchers themselves who may need to perform the QA, and arguably the results should be reported in research publications, along with validation experiments and uncertainty analysis.

The biggest challenge may be cultural. Much of MRI operates on a non-quantitative basis, and while there are fields with highly rigorous biophysical focus, it is rare to see a discussion of metrology in the context of verification and validation of a new technique or set of measurements, and the publication of measurement uncertainties is far from universal, even in otherwise-quantitative studies [168]. This can sometimes lead to subjective, qualitative concepts being applied in quantitative concepts. For example clinical recommendations for cardiac T1 mapping [169] require establishing scanner-specific reference values. In a fully quantitative context, we should instead be striving for universal reference values from well-defined metrological procedures. Culture is a more difficult thing to change than technical guidance, but it is perhaps best addressed by demonstrating what metrology and rigorously-reported measurements offer.

As with the metrological framework itself, if we build it, perhaps they will come.

Acknowledgements NPL's work on this manuscript was supported by the UK National Measurement System, funded by the Department of Science, Innovation, and Skills. The iMet-MRI project (20NRM05 iMet-MRI) has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme. Parts of the work shown here have received funding from the EMRP grant HLT06 MRI Safety. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union; from the EMPIR programme, co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme under grant number 17IND01 MIMAS; from the European Partnership of Metrology, co-financed by the Participating States and from the European Union's Horizon Europe research and innovation programme under grant number 21NRM05 STASIS and 18HLT05 QUIERO; from German Research Foundation (DFG), under grant number SCHM 2677/2-1; from the German Research Foundation (DFG) under the grant "BioQic" (No. GRK 2260) and from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme under grant number 22HLT02 A4IM. Parts of the work shown here was supported by the Korea Research Institute of Standard and Science under the project 'Development of medical data precision measurement technology' [KRIS-GP2024-0007-04]. NIST's work on this manuscript was supported by the National Institute of Standards and Technology, part of the United States Department of Commerce (<https://ror.org/05xpvk416>). The authors gratefully acknowledge the contribution of Paul Tofts at Brighton & Sussex Medical School, UK, and Alesandro Arduino, Orlando Bottauscio, Adriano Troia, and Umberto Zanollo of INRIM for useful discussions and advice. The authors would also like to thanks Emma Gates at NPL for providing technical review and feedback on the manuscript prior to submission.

Declarations

Conflict of interest Not applicable.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cercignani M, Dowell NG, Tofts PS (2018) Quantitative MRI of the brain: principles of physical measurement. CRC Press, USA
2. Seiberlich N, Gulani V, Campbell-Washburn A, Sourbron S, Calamante F, Hu HH (2020) Quantitative magnetic resonance imaging. Academic Press, USA

3. Martinez-Heras E, Grussu F, Prados F, Solana E, Llufríu S (2021) Diffusion-weighted imaging: recent advances and applications. *Semin Ultrasound CT MRI* 42(5):490–506. <https://doi.org/10.1053/j.sult.2021.07.006>
4. Keenan KE, Biller JR, Delfino JG, Boss MA, Does MD, Evelhoch JL, Griswold MA, Gunter JL, Hinks RS, Hoffman SW, Kim G, Lattanzi R, Li X, Marinelli L, Metzger GJ, Mukherjee P et al (2019) Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. *J. Magn. Reson. Imaging JMIR* 49:26–39. <https://doi.org/10.1002/jmri.26598>
5. Ruetten PPR, Gillard JH, Graves MJ (2019) Introduction to quantitative susceptibility mapping and susceptibility weighted imaging. *Br J Radiol.* <https://doi.org/10.1259/bjr.20181016>
6. Huber FA, Del Grande F, Rizzo S, Guglielmi G, Guggenberger R (2020) MRI in the assessment of adipose tissues and muscle composition: how to use it. *Quant Imaging Med Surg* 10(8):1636–1649. <https://doi.org/10.21037/qims.2020.02.06>
7. Ghassaban K, Liu S, Jiang C, Haacke EM (2019) Quantifying iron content in magnetic resonance imaging. *Neuroimage* 187:77–92. <https://doi.org/10.1016/j.neuroimage.2018.04.047>
8. Mezer A, Yeatman JD, Stikov N, Kay KN, Cho N-J, Dougherty RF, Perry ML, Parvizi J, Hua LH, Butts-Pauly K, Wandell BA (2013) Quantifying the local tissue volume and composition in individual brains with MRI. *Nat Med* 19(12):1667–1672. <https://doi.org/10.1038/nm.3390>
9. McRobbie DW, Moore EA, Graves MJ, Prince MR (2017) MRI from picture to proton. Cambridge University Press, Cambridge
10. Runge JH, Akkerman EM, Troelstra MA, Nederveen AJ, Beuers U, Stoker J (2016) Comparison of clinical MRI liver iron content measurements using signal intensity ratios, R2 and R2*. *Abdom Radiol* 41:2123–2131. <https://doi.org/10.1007/s00261-016-0831-7>
11. Kim BJ, Kang HG, Kim H-J, Ahn S-H, Kim NY, Warach S, Kang D-W (2014) Magnetic resonance imaging in acute ischemic stroke treatment. *J Stroke* 16(3):131–145. <https://doi.org/10.5853/jos.2014.16.3.131>
12. Albano D, Bruno F, Agostini A, Angileri SA, Benenati M, Bicchieri G, Cellina M, Chianca V, Cozzi D, Danti G, De Muzio F, Di Meglio L, Gentili F, Giacobbe G, Grazzini G, Grazzini I, Guerriero P, Messina C, Micci G, Palumbo P, Rocco MP, Grassi R, Miele V, Barile A (2022) Dynamic contrast-enhanced (DCE) imaging: state of the art and applications in whole-body imaging. *Jpn J Radiol* 40(4):341–366. <https://doi.org/10.1007/s11604-021-01223-4>
13. Weiskopf N, Edwards LJ, Helms G, Mohammadi S, Kirilina E (2021) Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nat Rev Phys* 3:570–588. <https://doi.org/10.1038/s42254-021-00326-1>
14. “UK Biobank.” [Online]. Available: <https://www.ukbiobank.ac.uk/>
15. J. J. Eertink, I. Bahce, J. C. Waterton, M. C. Huisman, R. Boellaard, A. Wunder, A. Thiele, and C. W. Menke van der Houven (2024) The development process of ‘fit-for-purpose’ imaging biomarkers to characterize the tumor microenvironment,” *Front. Med.*, <https://doi.org/10.3389/fmed.2024.1347267>.
16. Smeaton WA (2000) The foundation of the metric system in France in the 1790s. *Platin Met Rev* 44(3):125–134. <https://doi.org/10.1595/003214000X443125134>
17. Product Regulation and Metrology Bill [HL]: HL Bill 18 of 2024–25. 2024.
18. R. S. Endlaw, U.P. Petroleum Traders Association vs Union Of India & Ors, on 14 August, 2014. [Online]. Available: <https://indiankanoon.org/doc/115418138/>
19. Rodrigues Filho BA, Gonçalves RF (2015) Legal metrology, the economy and society: a systematic literature review. *Measurement* 69:155–163. <https://doi.org/10.1016/j.measurement.2015.03.028>
20. IEC, “International standards.” [Online]. Available: <https://www.iec.ch/publications/international-standards>
21. BIPM, “Measurement standards and metrological traceability (VIM).” [Online]. Available: https://www.bipm.org/documents/20126/54295284/VIM4_CD_210111c.pdf/a57419b7-790f-2cca-f7c9-25d54d049bf6
22. VIM, “International Vocabulary of Metrology,” Jan. 2021. [Online]. Available: https://www.bipm.org/documents/20126/54295284/VIM4_CD_210111c.pdf/a57419b7-790f-2cca-f7c9-25d54d049bf6
23. JCGM, “Evaluation of measurement data — Guide to the expression of uncertainty in measurement,” 2008. [Online]. Available: <https://www.bipm.org/en/committees/jc/jcgm/publications>
24. Yao A, Murbach M, Goren T, Zastrow E, Kainz W, Kuster N (2021) Induced radiofrequency fields in patients undergoing MR examinations: insights for risk assessment. *Phys Med Biol* 66(18):185014. <https://doi.org/10.1088/1361-6560/ac212d>
25. Winter L, Seifert F, Zilberti L (2021) MRI-Related heating of implants and devices: a review. *J Magn Reson Imaging* 53(6):1646–1665. <https://doi.org/10.1002/jmri.27194>
26. Arduino A, Bottauscio O, Bruhl R, Chiampi M, Zilberti L (2019) In silico evaluation of the thermal stress induced by MRI switched gradient fields in patients with metallic hip implant. *Phys Med Biol* 64(24):245006. <https://doi.org/10.1088/1361-6560/ab5428>
27. ISMRM, “Member-Initiated Sessions.” [Online]. Available: <https://www.ismr.org/25m/mis/>
28. JCGM, “International vocabulary of metrology—Basic and general concepts and associated terms (VIM),” version with minor corrections 2008. [Online]. Available: <https://www.bipm.org/en/committees/jc/jcgm/publications>
29. Losada-Urzaiz F, Gonzalez-Gaya C, Sebastian-Perez M (2015) Metrological regulations for quality control equipment calibration in pharmaceutical industry. *Procedia Eng* 132:811–815. <https://doi.org/10.1016/j.proeng.2015.12.564>
30. “History of Metrology,” MSC Training Symposium. [Online]. Available: <https://msc-conf.com/history-of-metrology/>
31. K. S. (Sammy) Li, “To Rule by Manufacture,” *Toung Pao*, vol. 103, no. 1–3, pp. 1–32, 2017.
32. “Standard measures of length,” Royal Museums Greenwich. [Online]. Available: <https://www.rmg.co.uk/collections/objects/rmgc-object-11210>
33. BIPM, “The International System of Units (SI).” [Online]. Available: <https://www.bipm.org/en/measurement-units#:~:text=From%2020%20May%202019%20all,technologies%2C%20to%20implement%20the%20definitions.>
34. Thomas RA, Bolt MA, Bass G, Nutbrown R, Chen T, Nisbet A, Clark CH (2017) Radiotherapy reference dose audit in the United Kingdom by the National Physical Laboratory: 20 years of consistency and improvements. *Phys Imaging Radiat Oncol* 3:21–27. <https://doi.org/10.1016/j.phro.2017.07.005>
35. Home Office, “Performance Standards and Test Protocols for Radiological Equipment.” [Online]. Available: <https://assets.publishing.service.gov.uk/media/5a74b9e1e5274a3cb2866b07/dose-rate-monitors.pdf>
36. “What is Cryptography,” IBM. [Online]. Available: <https://www.ibm.com/topics/cryptography>
37. NIST, “AI-Based Computational Metrology.” [Online]. Available: <https://www.nist.gov/mml/mmsd/materials-energy-and-sustainable-development-group/ai-based-computational-metrology>
38. National People’s Congress, “Metrology Law of the People’s Republic of China,” Law, Dec. 2018. [Online]. Available: <https://npcobserver.com/legislation/metrology-law/>

39. National Institute of Metrology China, “State Administration for Market Regulation: Regulation for National Metrology Specifications,” 2024. [Online]. Available: <https://en.nim.ac.cn/node/647>
40. “ISO/IEC 17025:2017.” [Online]. Available: <https://www.iso.org/standard/66912.html>
41. Analytical Methods Committee (2012) AMCTB No 53, Dark uncertainty. *Anal Methods* 4(9):2609–2612. <https://doi.org/10.1039/C2AY90034C>
42. Jones D (1994) ICRU report 50—prescribing, recording and reporting photon beam therapy. *Med Phys* 21(6):833–834. <https://doi.org/10.1118/1.597396>
43. Burakiewicz J, Christopher DJ, Sinclair DF, Walter GA, Kan HE, Hollingsworth KG (2017) Quantifying fat replacement of muscle by quantitative MRI in muscular dystrophy. *J Neurol* 264:2053–2067. <https://doi.org/10.1007/s00415-017-8547-3>
44. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, Huang W, Noworolski SM, Young RJ, Shiroishi MS (2019) Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials”. *J Magn Reson Imaging* 49(7):e101–e121. <https://doi.org/10.1002/jmri.26518>
45. Biomarkers Definitions Working Group (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69(3):89–95. <https://doi.org/10.1067/mcp.2001.113989>
46. Cashmore M, McCann A, Wastling SJ, McGrath C, Thornton J, Hall M (2021) Clinical quantitative MRI and the need for metrology. *Br J Radiol* 94:20201215. <https://doi.org/10.1259/bjr.20201215>
47. Reeder SB, Sirlin CB (2010) Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am* 18(3):337–357. <https://doi.org/10.1016/j.mric.2010.08.013>
48. Sullivan DA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, Marina Kondratovich L, McShane AP, Reeves DP, Barboriak AR (2015) Metrology standards for quantitative imaging biomarkers. *Radiology* 277:813–825. <https://doi.org/10.1148/radiol.2015142202>
49. European Union, “Directive 2013/35/EU of the European Parliament and of the Council of 26 June 2013 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (electromagnetic fields) (20th individual Directive within the meaning of Article 16(1) of Directive 89/391/EEC) and repealing Directive 2004/40/EC.” [Online]. Available: <https://eur-lex.europa.eu/eli/dir/2013/35/oj>
50. Chiampi M, Zilberti L (2011) Induction of electric field in human bodies moving near MRI: an efficient BEM computational procedure. *IEEE Trans Biomed Eng* 58(10):2787–2793. <https://doi.org/10.1109/TBME.2011.2158315>
51. Chiampi M, Zilberti L (2012) Reply to ‘Comments on induction of an electric field in human bodies moving near MRI: an efficient BEM computational procedure.’ *IEEE Trans Biomed Eng* 60(3):882–883. <https://doi.org/10.1109/TBME.2012.2232295>
52. Zilberti L, Chiampi M (2013) A numerical survey of motion-induced electric fields experienced by MRI operators. *Health Phys* 105(6):498–511. <https://doi.org/10.1097/HP.0b013e31829b4aac>
53. Zilberti L, Chiampi M, Bottauscio O (2015) Motion-induced fields in magnetic resonance imaging: are the dielectric currents really negligible? *IEEE Magn Lett* 6:1–4. <https://doi.org/10.1109/LMAG.2015.2429641>
54. Zilberti L, Bottauscio O, Chiampi M (2015) A potential-based formulation for motion-induced electric fields in MRI. *IEEE Trans Magn* 52(3):1–4. <https://doi.org/10.1109/TMAG.2015.2474748>
55. International Commission on Non-Ionizing Radiation Protection and others (2014) Guidelines for limiting exposure to electric fields induced by movement of the human body in a static magnetic field and by time-varying magnetic fields below 1 Hz. *Health Phys* 106(3):418–425. <https://doi.org/10.1097/HP.0b013e31829e5580>
56. Zilberti L, Bottauscio O, Chiampi M (2016) Assessment of exposure to MRI motion-induced fields based on the International Commission on Non-Ionizing Radiation Protection (ICNIRP) guidelines. *Magn Reson Med* 76(4):1291–1300. <https://doi.org/10.1002/mrm.26031>
57. International Electrotechnical Commission, “IEC 60601–2–33 Particular requirements for the basic safety and essential performance of magnetic resonance equipment for medical diagnosis,” 2022. [Online]. Available: <https://webstore.iec.ch/en/publication/67211>
58. Boulant N, Wu X, Adriany G, Schmitter S, Uğurbil K, Van De Moortele P (2016) Direct control of the temperature rise in parallel transmission by means of temperature virtual observation points: Simulations at 10.5 tesla. *Magn Reson Med* 75(1):249–256. <https://doi.org/10.1002/mrm.25637>
59. Eichfelder G, Gebhardt M (2011) Local specific absorption rate control for parallel transmission by virtual observation points. *Magn Reson Med* 66(5):1468–1476. <https://doi.org/10.1002/mrm.22927>
60. Graesslin I, Vernickel P, Börner P, Nehrke K, Mens G, Harvey P, Katscher U (2015) Comprehensive RF safety concept for parallel transmission MR. *Magn Reson Med* 74(2):589–598. <https://doi.org/10.1002/mrm.25425>
61. Guerin B, Angelone LM, Dougherty D, Wald LL (2020) Parallel transmission to reduce absorbed power around deep brain stimulation devices in MRI: Impact of number and arrangement of transmit channels. *Magn Reson Med* 83(1):299–311. <https://doi.org/10.1002/mrm.27905>
62. Padormo F, Beqiri A, Hajnal JV, Malik SJ (2016) Parallel transmission for ultrahigh-field imaging. *NMR Biomed* 29(9):1145–1161. <https://doi.org/10.1002/nbm.3313>
63. Petzold J, Schmitter S, Silemek B, Winter L, Speck O, Ittermann B, Seifert F (2023) Towards an integrated radiofrequency safety concept for implant carriers in MRI based on sensor-equipped implants and parallel transmission. *NMR Biomed* 36(7):e4900. <https://doi.org/10.1002/nbm.4900>
64. Petzold J, Schmitter S, Silemek B, Winter L, Speck O, Ittermann B, Seifert F (2023) Investigation of alternative RF power limit control methods for 0.5T, 1.5T, and 3T parallel transmission cardiac imaging: A simulation study. *Magn Reson Med*. <https://doi.org/10.1002/mrm.29932>
65. Petzold J, Schmitter S, Silemek B, Winter L, Speck O, Ittermann B, Seifert F (2024) Investigation of alternative RF power limit control methods for 0.5 T, 1.5 T, and 3T parallel transmission cardiac imaging: a simulation study. *Magn Reson Med* 91(4):1659–1675
66. Silemek B, Seifert F, Petzold J, Hoffmann W, Pfeiffer H, Speck O, Rose G, Ittermann B, Winter L (2022) Rapid safety assessment and mitigation of radiofrequency induced implant heating using small root mean square sensors and the sensor matrix Q_s . *Magn Reson Med* 87(1):509–527. <https://doi.org/10.1002/mrm.28968>
67. Silemek B, Seifert F, Petzold J, Brühl R, Ittermann B, Winter L (2023) Wirelessly interfacing sensor-equipped implants and MR scanners for improved safety and imaging. *Magn Reson Med* 90(6):2608–2626. <https://doi.org/10.1002/mrm.29818>
68. Winter L, Silemek B, Petzold J, Pfeiffer H, Hoffmann W, Seifert F, Ittermann B (2020) Parallel transmission medical implant safety testbed: real-time mitigation of RF induced tip

- heating using time-domain E-field sensors. *Magn Reson Med* 84(6):3468–3484. <https://doi.org/10.1002/mrm.28379>
69. Schoen N, Seifert F, Petzold J, Metzger GJ, Speck O, Ittermann B, Schmitter S (2022) The impact of respiratory motion on electromagnetic fields and specific absorption rate in cardiac imaging at 7T. *Magn Reson Med* 88(6):2645–2661. <https://doi.org/10.1002/mrm.29402>
 70. Clementi V, Zanovello U, Arduino A, Ancarani C, Baruffaldi F, Bordini B, Chiampi M, Zilberti L (2022) Classification scheme of heating risk during MRI scans on patients with orthopaedic prostheses. *Diagnostics*. <https://doi.org/10.3390/diagnostics12081873>
 71. Zilberti L, Bottauscio O, Chiampi M, Hand J, Lopez HS, Brühl R, Crozier S (2015) Numerical prediction of temperature elevation induced around metallic hip prostheses by traditional, split, and uniplanar gradient coils. *Magn Reson Med* 74(1):272–279. <https://doi.org/10.1002/mrm.25687>
 72. Zilberti L, Bottauscio O, Chiampi M, Hand J, Lopez HS, Crozier S (2014) Collateral thermal effect of MRI-LINAC gradient coils on metallic hip prostheses. *IEEE Trans Magn* 50(11):1–4. <https://doi.org/10.1109/TMAG.2014.2323119>
 73. Zilberti L, Arduino A, Bottauscio O, Chiampi M, Hand J, Lopez HS, Crozier S (2016) The underestimated role of gradient coils in MRI safety. *Magn Reson Med* 77(1):13–15. <https://doi.org/10.1002/mrm.26544>
 74. Arduino A, Bottauscio O, Chiampi M, Zanovello U, Zilberti L (2022) A contribution to MRI safety testing related to gradient-induced heating of medical devices. *Magn Reson Med* 88(2):930–944. <https://doi.org/10.1002/mrm.29235>
 75. Zanovello U, Fuss C, Arduino A, Bottauscio O (2023) Efficient prediction of MRI gradient-induced heating for guiding safety testing of conductive implants. *Magn Reson Med* 90(5):2011–2018. <https://doi.org/10.1002/mrm.29787>
 76. Arduino A, Zanovello U, Hand J, Zilberti L, Brühl R, Chiampi M, Bottauscio O (2021) Heating of hip joint implants in MRI: the combined effect of RF and switched-gradient fields. *Diagnostics* 85(6):3447–3462. <https://doi.org/10.1002/mrm.28666>
 77. Wooldridge J, Arduino A, Zilberti L, Zanovello U, Chiampi M, Clementi V, Bottauscio O (2021) Gradient coil and radiofrequency induced heating of orthopaedic implants in MRI: influencing factors. *Phys Med Biol*. <https://doi.org/10.1088/1361-6560/ac3eab>
 78. Zilberti L, Zanovello U, Arduino A, Bottauscio O, Chiampi M (2020) RF-induced heating of metallic implants simulated as PEC: is there something missing? *Magn Reson Med* 85(2):583–586. <https://doi.org/10.1002/mrm.28512>
 79. Arduino A, Baruffaldi F, Bottauscio O, Chiampi M, Martinez J, Zanovello U, Zilberti L (2022) Computational dosimetry in MRI in presence of hip, knee or shoulder implants: do we need accurate surgery models? *Phys Med Biol*. <https://doi.org/10.1088/1361-6560/aca5e6>
 80. Bottauscio O, Zanovello U, Arduino A, Zilberti L (2024) Polynomial chaos expansion of SAR and temperature increase variability in 3 T MRI due to stochastic input data. *Phys Med Biol*. <https://doi.org/10.1088/1361-6560/ad5070>
 81. Zilberti L, Arduino A, Torchio R, Zanovello U, Baruffaldi F, Lopez HS, Bettini P, Alotto P, Chiampi M, Bottauscio O (2023) Orthopedic implants affect the electric field induced by switching gradients in MRI. *Magn Reson Med* 91(1):398–412. <https://doi.org/10.1002/mrm.29861>
 82. Zilberti L, Curreli C, Arduino A, Zanovello U, Baruffaldi F, Bottauscio O (2024) Gradient-induced vibrations and motion-induced Lenz effects on conductive nonmagnetic orthopedic implants in MRI. *Magn Reson Med*. <https://doi.org/10.1002/mrm.30263>
 83. Zanovello U, Seifert F, Bottauscio O, Winter L, Zilberti L, Ittermann B (2022) CoSimPy: an open-source python library for MRI radiofrequency coil EM/circuit cosimulation. *Comput Methods Programs Biomed*. <https://doi.org/10.1016/j.cmpb.2022.106684>
 84. Bottauscio O, Chiampi M, Zilberti L (2014) Massively parallelized boundary element simulation of voxel-based human models exposed to MRI fields. *IEEE Trans Magn* 50(2):1029–1032. <https://doi.org/10.1109/TMAG.2013.2280523>
 85. Bottauscio O, Chiampi M, Zilberti L (2014) A hybrid FE–BE method for SAR estimate in voxel based human models undergoing MRI. *Eng Anal Bound Elem* 49:15–21. <https://doi.org/10.1016/j.enganabound.2014.04.014>
 86. Giordano D, Zilberti L, Borsero M, Chiampi M, Bottauscio O (2014) Experimental validation of MRI dosimetric simulations in phantoms including metallic objects. *IEEE Trans Magn* 50(11):1–4. <https://doi.org/10.1109/TMAG.2014.2323428>
 87. Giordano D, Zilberti L, Borsero M, Forastiere R, Wang W (2015) Validation of numerical methods for electromagnetic dosimetry through near-field measurements. *Acta Imeko* 4(1):90–96. https://doi.org/10.21014/acta_imeko.v4i1.169
 88. Zanovello U, Borsero M, Giordano D, Zilberti L, Maggiorelli F, Tiberi G (2017) Experimental setup to compare measurements and numerical simulations in magnetic resonance imaging RF dosimetry. *IEEE Trans Instrum Meas* 66(6):1208–1216. <https://doi.org/10.1109/TIM.2017.2687138>
 89. Arduino A, Bottauscio O, Chiampi M, Zilberti L (2017) Douglas–Gunn method applied to dosimetric assessment in magnetic resonance imaging. *IEEE Trans Magn*. <https://doi.org/10.1109/TMAG.2017.2658021>
 90. Bottauscio O, Chiampi M, Hand J, Zilberti L (2015) A GPU computational code for eddy-current problems in voxel-based anatomy. *IEEE Trans Magn*. <https://doi.org/10.1109/TMAG.2014.2363140>
 91. Bottauscio O, Cassarà A, Hand J, Giordano D, Zilberti L, Borsero M, Chiampi M, Weidemann G (2015) Assessment of computational tools for MRI RF dosimetry by comparison with measurements on a laboratory phantom. *Phys Med Biol*. <https://doi.org/10.1088/0031-9155/60/14/5655>
 92. U. Zanovello, L. Matekovits, and L. Zilberti, “An ideal dielectric coat to avoid prosthesis RF-artefacts in Magnetic Resonance Imaging,” *Sci. Rep.*, Mar. 2017, [Online]. Available: <https://www.nature.com/articles/s41598-017-00215-7>
 93. Zanovello U, Zilberti L, Matekovits L (2018) A near-field cloaking study to reduce mri rf-artefacts in presence of elongated prostheses. *J Electromagn RF Microw Med Biol* 2(4):249–256. <https://doi.org/10.1109/JERM.2018.2875248>
 94. Liu W, Wang H, Zhang P, Li C, Sun J, Chen Z, Xing S, Liang P, Wu T (2019) Statistical evaluation of radiofrequency exposure during magnetic resonant imaging: application of whole-body individual human model and body motion in the coil. *Int J Environ Res Public Health*. <https://doi.org/10.3390/ijerph16061069>
 95. speag, *SEMCAD*. [Online]. Available: <https://speag.swiss/products/semcad/overview/>
 96. National Electrical Manufacturers Association, *Characterization of the Specific Absorption Rate (SAR) for magnetic resonance imaging systems*, 2016. [Online]. Available: [https://www.nema.org/standards/view/Characterization-of-the-Specific-Absorption-Rate-for-Magnetic-Resonance-Imaging-Systems#:~:text=Resonance%20Imaging%20Systems-,Characterization%20of%20the%20Specific%20Absorption%20Rate%20\(SAR,for%20Magnetic%20Resonance%20Imaging%20Systems&text=Describes%20calorimetric%20and%20pulse%20energy,produce%20relatively%20homogeneous%20RF%20fields.](https://www.nema.org/standards/view/Characterization-of-the-Specific-Absorption-Rate-for-Magnetic-Resonance-Imaging-Systems#:~:text=Resonance%20Imaging%20Systems-,Characterization%20of%20the%20Specific%20Absorption%20Rate%20(SAR,for%20Magnetic%20Resonance%20Imaging%20Systems&text=Describes%20calorimetric%20and%20pulse%20energy,produce%20relatively%20homogeneous%20RF%20fields.)
 97. International Electrotechnical Commission, “IEC 62464–1:2019, Magnetic resonance equipment for medical imaging – Part 1:

- Determination of essential image quality parameters,” 2019. [Online]. Available: <https://webstore.iec.ch/en/publication/60565>
98. International Electrotechnical Commission, “IEC 60601–1:2005, Medical electrical equipment - Part 1: General requirements for basic safety and essential performance,” 2005. [Online]. Available: <https://webstore.iec.ch/en/publication/2606>
 99. Borraz O (2007) Governing standards: the rise of standardization processes in France and in the EU. *Governance* 20(1):57–84
 100. [Online]. Available: <https://openstd.samr.gov.cn/bzgk/gb/index>
 101. Siemens Medical, “Operator Manual: MAGNETOM Trio a Tim System.” 2007.
 102. Novikov DS, Kiselev VG, Jespersen SN (2018) On modeling. *Magn Reson Med* 79(6):3172–3193
 103. Waterman, *Iowa Med. Soc’y v. Iowa Bd. of Nursing*. 2013.
 104. American College of Radiology, “Large and medium phantom test guidance for the ACR MRI Accreditation Program (REVISED),” 19 2022. [Online]. Available: <https://www.acraccreditation.org/-/media/ACRAccreditation/Documents/MRI/ACR-Large--Med-Phantom-Guidance-102022.pdf>
 105. Huq MS, Fraass BA, Dunscombe PB, Gibbons JP Jr, Ibbott GS, Mundt AJ, Mutic S, Palta JR, Rath F, Thomadsen BR (2016) The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. *Med Phys* 43(7):4209–4262
 106. Institute of Physics & Engineering In Medicine, “Quality Control and Artefacts in Magnetic Resonance Imaging,” 112. Accessed: May 31, 2024. [Online]. Available: <https://www.ipem.ac.uk/resources/books/report-112-quality-control-and-artefacts-in-magnetic-resonance-imaging/>
 107. ICRU, *Determination of absorbed dose in a patient irradiated by beams of X or gamma rays in radiotherapy procedures ICRU report 24*. United States: International Commission on Radiation Units and Measurements, 1987. [Online]. Available: http://inis.iaea.org/search/search.aspx?orig_q=RN:25005992
 108. Ihalainen TM, Lonnroth NT, Peltonen JI, Uusi-Simola JK, Timonen MH, Kuusela LJ, Savolainen SE, Sipila OE (2011) MRI quality assurance using the ACR phantom in a multi-unit imaging center. *Acta Oncol* 50(6):966–972. <https://doi.org/10.3109/0284186X.2011.582515>
 109. Chien-Chuan C, Wan Y-L, Wai Y-Y, Liu H-L (2004) Quality assurance of clinical MRI scanners using ACR MRI phantom: preliminary results. *J Digit Imaging* 17(4):279–284. <https://doi.org/10.1007/s10278-004-1023-5>
 110. Weinreb J, Wilcox PA, Hayden J, Lewis R, Froelich J (2005) ACR MRI accreditation: yesterday, today, and tomorrow. *J Am Coll Radiol* 2(6):494–503. <https://doi.org/10.1016/j.jacr.2004.11.004>
 111. Clifford CR, Bernstein AM, Fox CN, Thompson P, Alexander G, Harvey D, Borowski B (2008) The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging*. <https://doi.org/10.1002/jmri.21049>
 112. Gunter JL, Bernstein MA, Borowski BJ, Ward CP, Britson PJ, Felmlee JP, Schuff N, Weiner M, Jack CR (2009) Measurement of MRI scanner performance with the ADNI phantom. *Med Phys*. <https://doi.org/10.1118/1.3116776>
 113. Wyman B, Harvey DJ, Crawford K (2012) Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers Dement*. <https://doi.org/10.1016/j.jalz.2012.06.004>
 114. American Association of Physicists in Medicine, “AAPM REPORT No. 100: Acceptance testing and quality assurance procedures for magnetic resonance imaging facilities,” College Park, MD: One Physics Ellipse, 2010.
 115. Stupic K, Ainslie M, Boss M, Charles C, Clarke L, Evelhoch J, Finn P, Hill D, Keenan K, Liu G, Yuan C, Martin M, Ainslie M, Rentz N (2021) A standard system phantom for magnetic resonance imaging. *Magn Reson Med*. <https://doi.org/10.1002/mrm.28779>
 116. M. Cashmore, C. Clarkson, B. P. Tatman, K. Obee, J. Clarke, S. Curtis, N. Smith, F. Brochu, E. Cooke, A. Ford-Scille, J. Goldring, R. Hanson, A. Ntata, S. Rhodes, S. Busoni, A. McCann, C. McGrath, R. Ferrero, A. Manzin, A. Troia, S. Hill, S. Rai, S. Strepokytov, C. Ward-Deitrich, A. Bosnjakovic, P. Tofts, T. Dispinar, I. Un, A. McDowell, S. Wastling, J. Thornton, N. Zafeiropoulos, S. Curtis, R. Scott, H. Elbert, J. Delve, C. Ingham, L. Pokvić Gurbeta, M. Smajlhodžić-Deljo, and M. Hall, “Enabling reproducible measurements of Fat Fraction and Iron content using an SI traceable reference phantom,” presented at the ISMRM & ISMRT Annual Meeting & Exhibition, 2024.
 117. Captur G, Bhandari A, Brühl R, Ittermann B, Keenan KE, Yang Y, Eames RJ, Benedetti G, Torlasco C, Ricketts L, Boubertakh R, Fatih N, Greenwood JP, Paulis LEM, Lawton CB, Bucciarrelli-Ducci C, Lamb HJ, Steeds R, Leung SW, Berry C, Valentin S, Flett A, De Lange C, DeCobelli F, Viallon M, Croisille P, Higgins DM, Greiser A, Pang W, Hamilton-Craig C, Strugnell WE, Dresselaers T, Barison A, Dawson D, Taylor AJ, Mongeon F-P, Plein S, Messroghli D, Al-Mallah M, Grieve SM, Lombardi M, Jang J, Salerno M, Chaturvedi N, Kellman P, Bluemke DA, Nezafat R, Gatehouse P, Moon JC (2020) T1 mapping performance and measurement repeatability: results from the multi-national T1 mapping standardization phantom program (TIMES). *J Cardiovasc Magn Reson* 22(1):31. <https://doi.org/10.1186/s12968-020-00613-3>
 118. Captur G, Gatehouse P, Keenan KE, Heslinga FG, Bruehl R, Prothmann M, Graves MJ, Eames RJ, Torlasco C, Benedetti G, Donovan J, Ittermann B, Boubertakh R, Bathgate A, Royet C, Pang W, Nezafat R, Salerno M, Kellman P, Moon JC (2016) A medical device-grade T1 and ECV phantom for global T1 mapping quality assurance—the T1 mapping and ECV standardization in cardiovascular magnetic resonance (TIMES) program. *J Cardiovasc Magn Reson* 18(1):58. <https://doi.org/10.1186/s12968-016-0280-z>
 119. Topriceanu C-C, Fornasiero M, Seo H, Webber M, Keenan KE, Stupic KF, Bruehl R, Ittermann B, Price K, McGrath L, Pang W, Hughes AD, Nezafat R, Kellman P, Pierce I, Moon JC, Captur G (2023) Developing a medical device-grade T2 phantom optimized for myocardial T2 mapping by cardiovascular magnetic resonance. *J Cardiovasc Magn Reson* 25(1):19. <https://doi.org/10.1186/s12968-023-00926-z>
 120. H.-M. Cho, C. Hong, C. Lee, H. Ding, T. Kim, and B. Ahn (2020) EGO-compatible modular mapping phantom for magnetic resonance imaging, <https://doi.org/10.1038/s41598-020-71279-1>.
 121. B. P. Tatman, R. Hanson, A. McDowell, E. Cooke, C. Clarkson, T. Dispinar, I. Un, S. Hill, S. Rai, A. Abukashabeh, A. McCann, C. McGrath, S. Curtis, H. Elbert, J. Delve, C. Ingham, S. Busoni, J. Clarke, J. Thornton, N. Zafeiropoulos, S. Wastling, A. Manzin, R. Ferrero, A. Troia, F. Brochu, A. Ford-Scille, J. Goldring, A. Ntata, K. Obee, S. Rhodes, M. Smajlhodžić-Deljo, A. Deumić, A. Bosnjakovic, P. Tofts, R. Scott, M. Cashmore, and M. Hall, “Inter-site and inter-scanner reproducibility across four qMRI measurands using SI traceable references,” presented at the Medical Image Understanding and Analysis, 2024. <https://doi.org/10.3389/978-2-8325-1244-9>.
 122. Salvati R, Hitti E, Bellanger J-J, Saint-Jalmes H, Gambarota G (2016) Fat ViP mri: virtual phantom magnetic resonance imaging of water–fat systems. *Magn Reson Imaging* 34(5):617–623
 123. Wülker C, Gessert NT, Doneva M, Kastrulin S, Ercan E, Nielsen T (2024) Digital reference objects for evaluating algorithm performance in MR image formation. *Magn Reson Imaging* 105:67–74. <https://doi.org/10.1016/j.mri.2023.10.008>

124. Pesapane F, Rotili A, Penco S, Nicosia L, Cassano E (2022) Digital twins in radiology. *J Clin Med* 11(21):6553. <https://doi.org/10.3390/jcm11216553>
125. Arduino A, Chiampi M, Zilberti L, Bottauscio O (2016) Alternative approaches to magnetic resonance-based electric properties tomography and local specific absorption rate estimation. *IEEE Trans Magn* 53(2):1–8
126. Arduino A, Zilberti L, Chiampi M, Bottauscio O (2017) CSI-EPT in presence of RF-shield for MR-coils. *IEEE Trans Med Imaging* 36(7):1396–1404
127. Arduino A, Chiampi M, Pennecchi F, Zilberti L, Bottauscio O (2017) Monte Carlo method for uncertainty propagation in magnetic resonance-based electric properties tomography. *IEEE Trans Magn* 53(11):1–4
128. Arduino A, Bottauscio O, Chiampi M, Zilberti L (2018) Magnetic resonance-based imaging of human electric properties with phaseless contrast source inversion. *Inverse Probl* 34(8):084002
129. A. Arduino, O. Bottauscio, M. Chiampi, and L. Zilberti, “Uncertainty propagation in phaseless electric properties tomography,” presented at the 2019 International Conference on Electromagnetics in Advanced Applications (ICEAA), IEEE, 2019, pp. 0412–0415.
130. D. Liang, J. Cheng, Z. Ke, and L. Ying, “Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks,” vol. 37, no. 1, pp. 141–151, 2020, <https://doi.org/10.1109/MSP.2019.2950557>.
131. Hammernik K, Schlemper J, Qin C, Duan J, Summers RM, Rueckert D (2021) Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination. *Magn Reson Med* 86(4):1859–1872. <https://doi.org/10.1002/mrm.28827>
132. Arduino A, Berta L, Coriasco M, Troia A, Zanovello U, Zilberti L, Morana G, Rampado O, Bottauscio O (2024) PS08. 10 development of an anthropomorphic brain phantom for magnetic resonance-based electric properties tomography. *Phys Med* 125:104203
133. Arduino A (2021) EPTlib: An open-source extensible collection of electric properties tomography techniques. *Appl Sci* 11(7):3237
134. Cencini M, Lancione M, Pasquariello R, Peretti L, Pirkel CM, Schulte RF, Buonincontri G, Arduino A, Zilberti L, Biagi L (2024) Fast high-resolution electric properties tomography using three-dimensional quantitative transient-state imaging-based water fraction estimation. *NMR Biomed* 37(1):e5039
135. Arduino A, Pennecchi F, Katscher U, Cox M, Zilberti L (2023) Repeatability and reproducibility uncertainty in magnetic resonance-based electric properties tomography of a homogeneous phantom. *Tomography* 9(1):420–435
136. Reeder SB, Hu HH, Sirlin CB (2012) Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. *J Magn Reson Imaging* 36(5):1011–1014. <https://doi.org/10.1002/jmri.23741>
137. Department of Radiology, Hacettepe University School of Medicine, Ankara, Turkey and I. S. Idilman (2022) Proton density fat fraction: magnetic resonance imaging applications beyond the liver. *Diagn Interv Radiol*. <https://doi.org/10.5152/dir.2021.21845>
138. Gofton C, Upendran Y, Zheng M-H, George J (2023) MAFLD: How is it different from NAFLD? *Clin Mol Hepatol* 29(Suppl):S17–S31. <https://doi.org/10.3350/cmh.2022.0367>
139. Nedrud MA, Chaudhry M, Middleton MS, Moylan CA, Lerebours R, Luo S, Farjat A, Guy C, Loomba R, Abdelmalek MF, Sirlin CB, Bashir MR (2023) MRI quantification of placebo effect in nonalcoholic steatohepatitis clinical trials. *Radiology* 306(3):e220743. <https://doi.org/10.1148/radiol.220743>
140. D. Hernando, S. D. Sharma, M. Aliyari Ghasabeh, B. D. Alvis, S. S. Arora, G. Hamilton, L. Pan, J. M. Shaffer, K. Sofue, N. M. Szeverenyi, E. B. Welch, Q. Yuan, M. R. Bashir, I. R. Kamel, M. J. Rice, C. B. Sirlin, T. Yokoo, and S. B. Reeder, “Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat-water phantom: Proton-Density Fat-Fraction Quantification at 1.5T and 3T,” *Magn. Reson. Med.*, 2017, <https://doi.org/10.1002/mrm.26228>.
141. Kim HJ, Cho HJ, Kim B, You M, Lee JH, Huh J, Kim JK (2019) Accuracy and precision of proton density fat fraction measurement across field strengths and scan intervals: a phantom and human study. *J Magn Reson Imaging* 50(1):305–314. <https://doi.org/10.1002/jmri.26575>
142. Hu HH, Yokoo T, Bashir MR, Sirlin CB, Hernando D, Malyarenko D, Chenevert TL, Smith MA, Serai SD, Middleton MS, Henderson WC, Hamilton G, Shaffer J, Shu Y, Tkach JA, Trout AT, Obuchowski N, Brittain JH, Jackson EF, Reeder SB (2021) Linearity and bias of proton density fat fraction as a quantitative imaging biomarker: a multicenter, multiplatform, multivendor phantom study. *Radiology*. <https://doi.org/10.1148/radiol.2021202912>
143. Navaratna R, Zhao R, Colgan TJ, Hu HH, Bydder M, Yokoo T, Bashir MR, Middleton MS, Serai SD, Malyarenko D, Chenevert T, Smith M, Henderson W, Hamilton G, Shu Y, Sirlin CB, Tkach JA, Trout AT, Brittain JH, Hernando D, Reeder SB (2021) and the RSNA Quantitative Imaging Biomarker Alliance—Proton Density Fat Fraction Biomarker Committee, “Temperature-corrected proton density fat fraction estimation using chemical shift-encoded MRI in phantoms.” *Magn Reson Med*. <https://doi.org/10.1002/mrm.28669>
144. Zhao R, Hernando D, Harris DT, Hinshaw LA, Li K, Ananthakrishnan L, Bashir MR, Duan X, Ghasabeh MA, Kamel IR, Lowry C, Mahesh M, Marin D, Miller J, Pickhardt PJ, Shaffer J, Yokoo T, Brittain JH, Reeder SB (2021) Multisite multivendor validation of a quantitative MRI and CT compatible fat phantom. *Med Phys* 48(8):4375–4386. <https://doi.org/10.1002/mp.15038>
145. Cho H-M, Hong C, Lee C, Ding H, Kim T, Ahn B (2020) LEGO-compatible modular mapping phantom for magnetic resonance imaging. *Sci Rep* 10(1):14755. <https://doi.org/10.1038/s41598-020-71279-1>
146. Lee S, Hwangbo S, Hong C, Lee C, Cho H-M (2024) Development and evaluation of emulsified reference materials for magnetic resonance imaging. *Metrologia*. <https://doi.org/10.1088/1681-7575/ad1f7f>
147. M. A. Boss, A. M. Dienstfrey, Z. Gimbutas, K. E. Keenan, J. D. Splett, K. F. Stupic, and S. E. Russek, “Magnetic Resonance Imaging Biomarker Calibration Service: Proton Spin Relaxation Times,” 250–97, p. 35, May 2018, <https://doi.org/10.6028/NIST.SP.250-97>.
148. S. E. Russek, K. Keenan, K. Stupic, N. Rentz, and M. Boss, “Magnetic Resonance Imaging Biomarker Calibration Service: NMR Measurement of Isotropic Water Diffusion Coefficient,” Mar. 2023, <https://doi.org/10.6028/NIST.SP.250-100>.
149. S. E. Russek, K. F. Stupic, and K. E. Keenan, “Chapter One - Traceability and uncertainty in NMR and MRI measurements,” vol. 111, pp. 1–68, 2024, <https://doi.org/10.1016/bs.arnmr.2024.06.001>.
150. Hall MG, Cashmore MT, McGrath C, McCann A, Tofts PS (2023) The perfect diagnostic imaging machine and what it means for quantitative MRI reproducibility. *IPEM-Transl* 6:100019
151. “KRISSE Sets Standard for MRI and CT Diagnosis of Fatty Liver,” Medical Metrology Group. [Online]. Available: https://www.krissre.kr/gallery.es?mid=a20803030300&bid=0023&b_list=

- 10&act=view&list_no=4737&nPage=1&vlist_no_npage=1&keyField=&orderby=
152. “Hyperfine Swoop,” Hyperfine Swoop. [Online]. Available: <https://hyperfine.io/>
 153. Chetcuti K, Chilingulo C, Goyal M, Vidal L, O’Brien N, Postels D, Seydel K, Taylor T (2022) Implementation of a low-field portable MRI scanner in a resource-constrained environment: our experience in Malawi. *Am J Neuroradiol* 43(5):670–674
 154. Coelho S, Baete SH, Lemberskiy G, Ades-Aron B, Barrol G, Veraart J, Novikov DS, Fieremans E (2022) Reproducibility of the standard model of diffusion in white matter on clinical MRI systems. *Neuroimage* 257:119290. <https://doi.org/10.1016/j.neuroimage.2022.119290>
 155. Zhu B, Liu J, Cauley S, Rosen BR, Rosen MS (2018) Image reconstruction by domain-transform manifold learning. *Nat Portf* 555:487–492. <https://doi.org/10.1038/nature25988>
 156. Zimmermann F, Kolbitsch C, Schuenke P, Kofler A (2024) PINQI: an end-to-end physics-informed approach to learned quantitative MRI reconstruction. *IEEE Trans Comput Imaging* 99:1–12. <https://doi.org/10.1109/TCI.2024.3388869>
 157. Liu F, Feng L, Kijowski R (2019) MANTIS: model-augmented neural network with incoherent k-space sampling for efficient MR parameter mapping. *Magn Reson Med* 82(1):174–188. <https://doi.org/10.1002/mrm.27707>
 158. Guo R, El-Rewaidy H, Assana S, Cai X, Amyar A, Chow K, Bi X, Yankama T, Cirillo J, Pierce P, Goddu B, Ngo L, Nezafat R (2022) Accelerated cardiac T1 mapping in four heartbeats with inline MyoMapNet: a deep learning-based T1 estimation approach. *J Magn Reson Cardiovasc*. <https://doi.org/10.1186/s12968-021-00834-0>
 159. Cohen O, Zhu B, Rosen M (2018) MR fingerprinting Deep RecOnstruction NEtwork (DRONE). *Magn Reson Med*. <https://doi.org/10.1002/mrm.27198>
 160. Afroogh S, Akbari A, Malone E, Kargar M, Alambeigi H (2024) Trust in AI: progress, challenges, and future directions. *Humanit Soc Sci Commun* 11(1):1568. <https://doi.org/10.1057/s41599-024-04044-8>
 161. Shimron E, Perlman O (2023) AI in MRI: computational frameworks for a faster, optimized, and automated imaging workflow. *Bioengineering* 10(4):492. <https://doi.org/10.3390/bioengineering10040492>
 162. Lachin JM (2004) The role of measurement reliability in clinical trials. *Clin Trials*. <https://doi.org/10.1191/1740774504cn0570a>
 163. Curtis L. Meinert, *Clinical Trials: Design, Conduct and Analysis*. Oxford University Press, 1986. [Online]. Available: <https://doi.org/10.1093/acprof:oso/9780195035681.001.0001>
 164. Tofts PS (2022) The perfect qMR machine: measurement variance much less than biological variance. *Phys Med* 104:145–148. <https://doi.org/10.1016/j.ejmp.2022.10.013>
 165. Layton K, Kroboth S, Jia F, Littin S (2017) Pulseseq: a rapid and hardware-independent pulse sequence prototyping framework. *Magn Reson Imaging*. <https://doi.org/10.1002/mrm.26235>
 166. Karakuzu A, Boudreau M, Duval T, Boshkovski T, Leppert I, Cabana J-F, Gagnon I, Beliveau P, Pike GB, Cohen-Adad J, Stikov N (2020) qMRLab: quantitative MRI analysis, under one umbrella. *J Open Source Softw*. <https://doi.org/10.21105/joss.02343>
 167. Karakuzu A, Biswas L, Cohen-Adad J, Stikov N (2022) Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. *Magn Reson Med* 88(3):1212–1228. <https://doi.org/10.1002/mrm.29292>
 168. Fokkinga E, Hernandez-Tamames J, Ianus A, Nilsson M, Tax CMW, Perez-Lopez R, Grussu F (2023) Advanced diffusion-weighted mri for cancer microstructure assessment in body imaging, and its relationship with histology. *J Magn Reson Imaging*. <https://doi.org/10.1002/jmri.29144>
 169. Messroghli DR, Moon JC, Ferreira VM, Grosse-Wortmann L, He T, Kellman P, Mascherbauer J, Nezafat R, Salerno M, Schelbert EB, Taylor AJ, Thompson R, Ugander M, Van Heeswijk RB, Friedrich MG (2016) 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* 19(1):75. <https://doi.org/10.1186/s12968-017-0389-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.