BJR https://doi.org/10.1259/bjr.20201215

Received: 09 October 2020 | Revised: 17 December 2020 | Accepted: 17 December 2020

Cite this article as:
Cashmore MT, McCann AJ, Wastling SJ, McGrath C, Thornton J, Hall MG. Clinical quantitative MRI and the need for metrology. *Br J Radiol* 2021; **94**: 20201215.

REVIEW ARTICLE

Clinical quantitative MRI and the need for metrology

[1]MATT T CASHMORE, PhD, [2]AARON J MCCANN, PhD, [3]STEPHEN J WASTLING, PhD, [2]CORMAC MCGRATH, PhD, [3]JOHN THORNTON, PhD and [1]MATT G HALL, PhD

[1]National Physical Laboratory, Teddington, UK
[2]Belfast Health and Social Care Trust, Belfast, Ireland
[3]Neuroradiological Academic Unit, UCL Institute of Neurology, University College London, London, UK

Address correspondence to: Dr Matt T Cashmore
E-mail: *matt.cashmore@npl.co.uk*

ABSTRACT

MRI has been an essential diagnostic tool in healthcare for several decades. It offers unique insights into most tissues without the need for ionising radiation. Historically, MRI has been predominantly used qualitatively, images are formed to allow visual discrimination of tissues types and pathologies, rather than providing quantitative measurements. Increasingly, quantitative MRI (qMRI) is also finding clinical application, where images provide the basis for physical measurements of, *e.g.* tissue volume measures and represent aspects of tissue composition and microstructure. This article reviews some common current research and clinical applications of qMRI from the perspective of measurement science. qMRI not only offers additional information for radiologists, but also the opportunity for improved harmonisation and calibration between scanners and as such it is well-suited to large-scale investigations such as clinical trials and longitudinal studies. Realising these benefits, however, presents a new kind of technical challenge to MRI practioners. When measuring a parameter quantitatively, it is crucial that the reliability and reproducibility of the technique are well understood. Strictly speaking, a numerical result of a measurement is meaningless unless it is accompanied by a description of the associated measurement uncertainty. It is therefore necessary to produce not just estimates of physical properties in a quantitative image, but also their associated uncertainties. As the process of determining a physical property from the raw MR signal is complicated and multistep, estimation of uncertainty is challenging and there are many aspects of the MRI process that require validation. With the clinical implementation of qMRI techniques and its continued expansion, there is a clear and urgent need for metrology in this field.

INTRODUCTION

In recent years, there has been increasing development and use of quantitative imaging techniques in health care. This move towards increased quantitation requires a corresponding increase in both validation and support. Like any other measurement, the bias, uncertainty and reproducibility of the quantities derived from imaging need to have their limitations quantified and assessed in order to use them with confidence. This is particularly important if they are used to inform clinical decisions. The science of metrology provides the framework by which we can perform this assessment. This is a well-established principle – quantitative measurements are used extensively in daily life, be they length, mass or time, and are fundamental to modern society as they underpin core concepts such as GPS, economics and manufacturing. To ensure confidence in these measurements, we need to understand the levels of variance and uncertainty associated with them. Measurables such as length are made with respect to a known reference, and along with descriptions of measurement uncertainty give a means of describing confidence in a measurement process. As quantitative imaging continues to grow and develop, it is necessary that we begin to incorporate metrology into general use, so that we understand how to use these values to maximise their effectiveness whilst minimising the risk of error. This paper will review quantitative MRI through the lens of metrology and will highlight the importance of measurement science in this growing field.

Quantitative MRI

MRI is an essential tool for examining tissue structure and function, enabling the diagnosis and monitoring of many conditions and pathologies. It is also an important tool in the development of new therapies and has applications in treatment planning. Many established and emerging applications require quantitative estimates – *measurements* – of some physical property, *e.g.* tumour volume or liver iron

concentration. In any measurement process, it is important to understand not just the measured values but also their associated uncertainties. Without this, values cannot be meaningfully compared since there is no notion of the significance of differences in value.

Image formation and the estimation of parameters from image data is a non-trivial and multistep process in MRI. Both acquisition and processing are potential sources of bias and uncertainty. This means that to meaningfully make use of qMRI data, careful calibration and testing of MRI instruments and associated image analysis tools is critical, in a similar vein to that currently employed in imaging modalities which use ionising radiation. Users of MRI systems need to be confident that parameters derived from quantitative imaging techniques are reliable and robust. Ensuring metrological rigour at each stage of the measurement chain, *e.g.* test object calibration or software verification, increases trust in the data being used to inform clinical decisions.

The current fleet of MRI systems in the UK is highly diverse. Many institutions have a range of machines varying in age, model, manufacturer, and field strength. Most are 1.5 and 3 T systems with a 60 or 70 cm bore. The dominant manufacturers are GE Healthcare, Philips Healthcare, and Siemens Healthineers although other firms such as Canon Medical are also active. Pre-clinical, narrow bore systems with similarly diverse specifications and age ranges are also common. Main fields on these systems vary from 4.7 to 11.4 T. There is also a movement towards specialist MR scanners with unusual form factors such as upright and portable units and a growing number of MR-Linac systems and PET-MR facilities.

Most MRI procedures currently requested by physicians are qualitative. Image contrast is relative, and no quantitative parameters are derived from the images – diagnostic decisions are based on the interpretation of a radiologist. Images of this kind are highly useful for informing clinical decisions about individual patients but are problematic for making comparisons between groups of patients or sets of timepoints – there are no reported uncertainties, and different scanners give different intensities. Moreover, scanners may experience drift over time and experience step-changes after servicing or upgrades.

Quantitative measurement processes allow for MR scanning to infer parameters which can be traced to verifiable physical quantities, thus potentially improving consistency. This makes it more practical to perform large-scale studies, as images can be processed and analysed according to the same criteria. Coupled with quantified measurement uncertainty, this is potentially transformative for larger trials and for training artificial intelligence (AI) approaches. The techniques of quantitative MRI cover a diverse range of approaches and there has been substantial work elsewhere discussing the biological origins and specific MRI considerations relevant to this area. See Cercignani et al[1] for an in-depth technical overview.

Metrology

Metrology is the study of measurement processes. The Bureau International des Poids et Mesures (BIPM) defines metrology as "the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology".[2] This is realised through the key idea of *Traceability*; the chain of comparisons which directly relates any given measurement to the primary standard determination of that unit (*e.g.* metre, second, etc).

Without understanding the traceability of a measurement, there is no way of making a meaningful comparison between values. Without an evaluation of the measurement uncertainty, we have no way of knowing whether a difference is significant. Within health care, through the involvement of National Measurement Institutes (NMIs) in the field of radiotherapy, clinical practitioners can be confident that the dose delivered to a patient in one hospital can be directly compared to the dose delivered elsewhere in the country. All systems are calibrated via a single traceability chain to a common primary standard. There is currently no equivalent level of data confidence available in MRI, but quantifying the relationship between the "true" value of a measurand, and one determined through MR techniques is a rapidly developing field of study.

Recent published work has highlighted the importance of metrology in health care.[3,4] Developing and including this level of rigour in qMRI requires more education and awareness of concepts such as traceability. Understanding how uncertainty can propagate through the measurement process is vital. Without this, seemingly trivial variables may be ignored or dismissed as their impact on the desired result is underestimated. The BIPM define a quantity as: "a property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference". The BIPM define reference data as being "Data related to a property of a phenomenon, body, or substance, or to a system of components of known composition or structure, obtained from an identified source, critically evaluated and verified for accuracy".[2] Within qMRI, the "number" is the measurement as determined from the scanner, however metrology provides the capability to define the "reference" with which that measurement is made. Uncertainties in equipment used to measure a sample in a phantom will, *e.g.* contribute to the overall uncertainty in measurement of that test object. It is also important to recognise that a quantified measurement uncertainty allows for confidence in data and is not merely describing an error in its quantity.

Quantitative biomarkers

A biomarker is a biological characteristic that can be objectively measured as an indicator of normal or pathogenic biological processes, or pharmacological responses to therapeutic intervention.[5–7] Biomarkers can be measured in a variety of ways. Some examples in clinical use include urine peptide content for monitoring kidney disease,[8] blood-based microRNA profiling for early diagnosis of pancreatic cancer,[9] those involved in the RECIST system for tumour response,[10] and the presence of circulating tumour cells (CTCs) in biopsies.[11] Biomarkers such as the presence of CTCs have been the subject of significant study, with a well-understood biological mechanism linking their presence to the risk of metastasis in a variety of cancers.[12,13]

A quantitative *imaging* biomarker (QIB) is one which is extracted specifically from a quantitative imaging technique, *i.e.* one where it is possible to quantitatively measure properties from a given image. One advantage of image-based biomarkers is their wide-field coverage of a tissue of interest. This is not possible in more traditional assays such as biopsy. MRI potentially provides access to a wealth of imaging biomarkers, often with less patient risk than other methods.

Some common quantitative measurements with application as biomarkers in MRI are:

- T1, T2 and T2* relaxation times, *e.g.* measuring tissue oxygenation.[14]
- Apparent diffusion coefficient (ADC), *e.g.* detecting ischaemic stroke.[15]
- Tissue fat fraction, *e.g.* investigating muscular dystrophy.[16]
- Tissue iron fraction, *e.g.* diagnosing iron overload.[17]

Any biomarker used in clinical decision-making requires associated clinical guidelines. To use these reliably and safely, it is important that measurements made are presented with their corresponding uncertainty. Borderline cases where results lie close to diagnostic thresholds may be resolved through consideration of how the result and its uncertainty are related. Improper adoption of an insufficiently validated biomarker can result in endangering patient lives, as seen in the Cardiac Arrhythmia Suppression Trial,[18] highlighting how important it is to have well-verified measurement techniques and uncertainty analysis.

It is critical for both patient management and translational research applications that QIB measurements from different scanner installations are comparable. This is particularly true in multisite trials, which rely on the compatibility of data between sites and all participants. An inability to compare results obtained on different scanners may restrict the cohort of scanners in a trial and limit the utility of its findings.

The simplest form of a clinical trial tests a new therapy by comparing a treatment group and a control group. The statistical power of the study is critical to demonstrating a therapeutic effect, but where effect sizes are small, they are easily hidden by variability within groups. The conventional approach to solving this problem is to increase the sample size, but this increases costs and in the cases of rarer conditions may make recruitment impossible. By decreasing intersite variation, the statistical power of a trial increases without increasing sample size and hence costs. Quantifying uncertainties and bias in the measurement process makes it possible to calibrate them to minimise intersite variability and maximise statistical power.

There are currently two major international initiatives to incorporate quantitative imaging biomarkers into clinical practice: the US-led Quantitative Imaging Biomarkers Alliance (QIBA)[19] and the European Imaging Biomarkers Alliance (EIBALL).[20] Both consortia work with industry, research institutions, and healthcare providers to establish standards on which to base quantitative image-based measurements. QIBA defines measurement profiles which recommend guidelines for the measurement of quantitative biomarkers. NMIs are also increasingly becoming involved in work to support quantitative MRI biomarkers, as they provide the capability to traceably relate clinical measurands to fundamental primary standards. These include the National Physical Laboratory in the UK, the National Institute for Standards and Technology in the US, Physikalisch-Technische Bundesanstalt in Germany, Istituto Nazionale di Ricerca Metrologica in Italy, and the Korea Research Institute of Standards and Science. In recent years, the UK Biobank has also been leading a major effort in obtaining quantitative MRI scans of approximately 10,000 patients, to create a data set resource suitable for the development of imaging standards.[21]

Clinically available quantitative MRI

This section reviews some techniques used in clinical applications of quantitative MRI and applications in measuring intensive parameters.

Relaxometry

Relaxometry is one of the most common measurements in MRI. T1 and T2 relaxation times are the recovery and decay time constants of longitudinal and transverse magnetisation, respectively. Many clinical scans are optimised to reveal high contrast between tissues with different relaxation times. This allows radiologists to differentiate, *e.g.* lesion from healthy tissue, or white matter from grey in the brain.

A qualitative acquisition, such as a T_2-weighted image, provides tissue contrast by allowing the signals to decay for a specified time and then capturing the signal intensity for each voxel. Tissues with longer T2 will therefore appear brighter than those with shorter. The actual signal intensities in the image will be complex functions of the hardware and software employed to generate them. Different scanners will form images which are visually similar but, when viewed quantitatively, will exhibit considerable variation between sites and time points.

A quantitative image, such as a T2 *map*, has intensities set via estimates of the T2 parameter itself in each scan voxel. T2 maps are formed by sampling the observed signal decay curve at several different times and performing additional analysis to estimate the parameter of interest. Whilst the individual image intensities measured may be quite different from scanner to scanner, the relaxation behaviour of the tissue magnetisation state, as a function of applied magnetic field and time, is not. This makes images from different systems directly comparable, and provides new information not present in qualitative T2 images.[22]

Changes in the relaxation behaviour of tissue can be indicative of pathology. One example of this is multiple sclerosis (MS), where the lipid-rich myelin sheath surrounding neurons is degraded and the signal from myelin replaced by that from free water.[23] In qualitative MRI, a lesion may appear hypointense on a T_1 weighted image and hyperintense with T_2 weighting. Quantitative relaxation time mapping allows for more detailed understanding of underlying pathology and allow for improved prognostic estimation.[24] Quantitative relaxometry can also reduce the need for potentially harmful gadolinium-based contrast agents (GBCA).

Measurements of T1 in brain have also been shown to predict areas of contrast enhancement in glioblastoma patients.[25]

Liver iron measurements

One conspicuous application of relaxometry is the measurement of liver iron concentration (LIC). This is used in the diagnosis of iron overload (haemochromatosis), a systemic disorder which can lead to cirrhosis, cancer[26] and cardiac disease.[27] As the liver is one of the primary sites for storing iron in the body, it is amongst the first organs to show iron overload. Historically, liver biopsy followed by histology was the gold-standard for diagnosis; however, this is an invasive procedure and involves qualitative visual assessment of cell stains. Detection of elevated iron in the liver can allow earlier treatment of haemochromatosis and minimise permanent liver damage.

Iron storage primarily involves the protein ferritin, and the complex haemosiderin found within cells.[28] These two molecules alter the T2 and T2* of the liver parenchyma, allowing for their detection and quantification using relaxometry techniques. Qualitative detection is possible using fast spin echo sequences for T_2-weighted imaging, and single echo gradient echo (GRE) sequences for T2*-weighted imaging[29,30] but these approaches cannot accurately assess the degree of iron overload present and may not be a suitable basis for treatment. Quantitative measures of LIC allow for more accurate and precise staging of the disease and consequentially more informed treatment.

The most common method of quantifying LIC is by directly measuring the rate of signal decay associated with T2 relaxation, known as R2, and defined as

$$R2 = 1/T2 \quad (1)$$

This technique has received FDA approval[31] and is distributed commercially under the name Ferriscan.[32] R2 is used as it results in decreased sensitivity to noise, and also provides an easy-to-fit model relating it to iron concentration.[33,34] R2 is measured by fitting a signal curve to multiecho data, and a known relationship between R2 and LIC[35] further allows quantification of liver iron. R2 measurement is currently considered the gold-standard in MR measurement of LIC.

Fat fraction

The proportion of fat in tissue is also an important image-based biomarker. It is known as proton density fat fraction (PDFF).[36] Non-alcoholic fatty liver disease (NAFLD) is a pathology characterised by excessive fat deposition in liver cells and is one of the most common causes of liver disease.[37] Historically the gold-standard for diagnosis of NAFLD was liver biopsy followed by histology.[38] This is an invasive procedure and carries significant risk of complications.[39] It can also misrepresent liver health due to the heterogenous nature of NAFLD.

PDFF can be measured quantitatively with MRI.[38] The different chemical environment for protons in fat *vs* those in water leads to different resonant frequencies in response to an applied field. Consequentially, the MR signal detected from protons in fat shows a frequency shift from those in water and a different echo time. By careful selection of echo times, the water and fat signals can be measured in-phase (IP) and out-of-phase (OP). It is then possible to perform fat quantification based on the differences between signals. This is Dixon Imaging.[40]

In its original form (now known as 2-point), Dixon imaging produces fat and water images through the sum and difference of IP and OP images respectively. However, this version of the technique assumes that the dominant component of the signal is water, and as such can lead to incorrect results. A modified Dixon approach uses three images taken at different values of phase difference provides additional information required to resolve the ambiguity.[41] Unlike CT or ultrasound, MRI-based methods of investigating hepatic fat, show a much higher degree of sensitivity to steatosis, and allows for the differentiation of steatosis from iron overload.[42]

Diffusion

Applying additional pulsed field gradients during the MR pulse sequence allows signals to be sensitised to the bulk incoherent motion of spins. This is known as diffusion-weighted imaging (DWI) in which the measured quantity is the ADC. DWI is widely used in the examination of stroke.[43] Early stage ischaemic infarction in the brain is associated with cytotoxic oedema.[44] As the extracellular volume is reduced due to swelling, there is a corresponding reduction in the observed ADC using DWI.[45] This is reversed as the infarction progresses through the subacute stage towards chronic stroke due to lysis of the brain matter, increasing extracellular volume and allowing more isotropic diffusion of water resulting in high ADC. Diffusion MRI is a vital tool in the staging and monitoring of cerebral ischaemic stroke.

A more sophisticated application of diffusion imaging involves acquiring a set of images with gradients applied in different directions. This allows orientational differences in ADC to be measured. Measurements of directional diffusion are often fitted using a tensor, a technique is known as diffusion tensor imaging (DTI)[46] or more detailed and flexible models (HARDI).[47] By extracting the orientations in which the ADC is maximised, both DTI and HARDI provide estimates of the orientation of the underlying tissue. This has applications in fibrous tissue, such as white matter in the brain.[48] Tractography is a term used to describe a range of techniques which use DWI to non-invasively reconstruct white matter structures in the brain. The ability to perform *in vivo* mapping of brain structure is highly beneficial to neurosurgery, allowing for pre-operative planning as well as providing the capacity for intraoperative neuronavigation.[49]

DWI can also be applied in monitoring of post-treatment tumour growth. Here, it is important to differentiate between actual tumour recurrence, and non-neoplastic changes in tissue caused by radiation therapy (commonly called pseudoprogression).[50] The diffusion microenvironment experienced by spins in regions of tightly packed cell growth, such as tumours, is different than in regions of necrotic tissue such as those caused by radiation treatment. This leads to changes in the shapes and sizes of diffusion tensors which can be used to inform treatment plans and monitor patient response.[51]

Commercial software solutions

MRI manufacturers and other third parties provide a range of quantitative measurement packages. These are proprietary, unique to each manufacturer, and often include specific pulse sequences and software tools. There are a wide range of packages covering different biomarkers. Notably, Ferriscan (Resonance Health) is an FDA-approved service for measuring liver iron. There are also a variety of packages capable of measuring fat-fraction in tissue including:

- LiverLab, Siemens Healthineers.
- mDIXON Quant, Philips Healthcare.
- IDEAL IQ, GE Healthcare.

Despite their widespread usage, these software packages represent a new challenge in terms of analysing measurement uncertainty. As the algorithms used in processing data are proprietary and not visible to the user, there is no way to be able to accurately model uncertainty propagation through them. This is not a challenge unique to these examples, however, but applies also to the software solutions such as those present in the scanner systems. Data analysis routines will in general introduce a new uncertainty contribution and being able to model and understand these are an active topic of research in qMRI metrology.

DISCUSSION

There are many more applications of quantitative MRI which are used in the clinic and research. The techniques outlined here represent a survey of the most common, a comprehensive survey would cover several papers this length. At the time of writing, quantitative techniques clinically remain niche, but are certainly growing. Being able to extract quantitative information from imaging techniques allows for a wide range of analysis to be performed. Furthermore, the removal of subjective qualitative determinations, allows for a more rigorous comparison of scanner performance, a more detailed understanding of the pathology and biomechanics of disease progression, and consequently improved diagnostic and therapeutic capabilities.

Test objects

Calibration is a crucial aspect of any measurement process. To reliably compare two measurements, there must be a common point of reference between them. In MRI, this conventionally comes from scanning a test object, referred to as a phantom, which has known, well-characterised behaviours and (ideally) is traceable to a primary standard. It is not possible to calibrate an MR scanner solely from data acquired of humans, as their tissues are highly complex with material properties that cannot be sufficiently characterised for use as a reference.

MRI enables a range of different measurements, and as such a wide range of phantoms are necessary. Relaxometry phantoms typically consist of a gel or solution, doped with ionic salts to alter their relaxation times.[52,53] For measurements of image distortion, a phantom consisting of an array of MR-invisible plastic vials filled with appropriate contrast generating media is typically used. Frequently (but not universally), the object is also flood-filled with water.[54] An isotropic flood-filled phantom can be used for investigations of signal-to-noise ratio (SNR). There are also so-called system phantoms which are combination test objects designed to characterise a range of quantities.[40,41]

It is important that the phantom itself be routinely monitored for changes in its internal structure or chemical properties. A phantom can undergo a variety of chemical and physical changes over its lifetime, and data collected at different times may not be comparable. Organic gels, if prepared incorrectly, can show fungal growth. Imperfectly sealed compartments may result in leaking or evaporation. Plastics may be porous and subject to warping. A phantom should be monitored for changes over its lifetime to enable it to remain a useful reference.

Phantom cost is a significant barrier to uptake in clinical environments. MRI is not subject to the same legislative controls as CT or positron emission tomography (PET). There, legislation mandates a testing programme which optimises image quality and radiation dose. In the absence of strict regulatory justification, it is more difficult for financially constrained organisations to justify the purchase and maintenance of phantoms. One common mechanism to obtain phantoms is by participating in clinical trials. However, as different trials make use of different test objects, this leads to inconsistencies in available test objects at different sites. This disparity complicates the development of small-scale collaborative research as there are often difficulties in comparing measurements.

An alternative strategy for increasing the availability of effective Quality Assurance (QA) is to perform external audits. Rather than requiring all sites to perform their own procedures using their own phantoms, an external team performs the audit as a service. A small number of test objects may then be sufficient to provide traceable metrology to many scanners. Phantoms used for regular QA may not need to be as highly characterised as those used externally and may be periodically calibrated against them. The introduction of a chain of traceability can widen the availability of high-quality reference metrology where cost may otherwise be a barrier to uptake.

Some trial phantoms are relatively common. The Alzheimers Disease Neuroimaging Initiative (ADNI) is an ongoing large multicentre study in which all participating sites have their own ADNI phantom. It contains components suitable for measuring SNR, CNR and geometric distortion. ADNI was the first large-scale MRI trial to implement phantom scans by all centres. It revealed the presence of anomalies which would otherwise have remained undetected. Gradient stability was shown to be disrupted by scanner software and hardware upgrades, and the effect of servicing and upgrades was observed to be more significant than long-term drift.[55] ADNI found that by including longitudinal phantom scanning they discovered system errors such as misidentification of gradient hardware, disabling of automated-shimming and miscalibrations in positioning laser alignment. Abnormalities were found in over 25% of enrolling sites and if undetected would have led to aberrant study results.

NIST has developed a suite of quantitative measurement phantoms for use in MRI which include diffusion, breast and system

phantoms. The NIST System Standard Model 130 phantom allows for reference measurements of T1, T2, and proton density through a set of spheres containing water (regular and deuterated), doped with paramagnetic ion solutions.[56] The concentrations of these solutions are traceably verified to allow a comparison of interphantom variability. The relaxation properties of the test spheres are designed to cover a clinically relevant range of tissue behaviour. The phantom also contains inserts which can be used to measure slice thickness as well as spatial resolution, and an array of spheres for characterising three-dimensional distortion. NIST provides a series of recommended imaging protocols and has also produced software for analysing results. Thus, eliminating estimate variations arising from differing analysis implementations.

Limitations of this phantom include needing to account for the temperature dependence of the relaxometry components as well as difficulty in positional repeatability between different scans. Due to the nature of a system phantom, samples need to be compactly arranged, this can necessitate the use of sequences which take more time than is practical in a clinical setting. The spatial resolution inset requires qualitative inspection and presents difficulties for objective, automated measurements. As the NIST system phantom covers a wide range of measurables, it is relatively costly, and may not be affordable to many scanning facilities.

MRI as a measurement device

As a medical device, MR scanners fall under the remit of the international standard IEC 60601. Crucially for the development of quantitative techniques, however, this standard does not recognise MR scanners as a device with a measurement capability. Partially because of this, there is an ambiguity in international standardisation covering the use of quantitative MR-based techniques.

In the UK, the MHRA provides guidance on the safe use of clinical MRI.[57] This guidance outlines a basic degree of quality assurance for MRI equipment and reinforces the need for regular quality testing. Also included is an emphasis on the need for performing geometric distortion and signal scaling verifications, as well as the necessity of having well-characterised test objects for this purpose. IPEM advises that "It is important that the images and data produced by the MRI system are accurate, reliable, and representative of the patient and of high diagnostic quality. An appropriate QA program should be adopted to ensure that the MRI system continues to meet a standard for diagnostic use through system performance monitoring, audit, review and any necessary corrective action.".[58] Whilst these commentaries acknowledge the value of QA in MRI procedures, they do not formulate a code of practice for repeatable results with suggestions on specific test objects, conditions or sequences.

The Magnetic Resonance National Evaluation Team (MagNET) was a UK-based non-profit organisation that offered type testing and performance evaluations of MRI systems from the late 1980s through to the early 2000s. Reports from MagNET were used extensively in the NHS for purchasing and acceptance-testing during this time.[59] MagNET visited manufacturing facilities to perform image performance tests on MRI hardware and published test reports for each model. These provided end-users with an independent assessment of scanner capability which helped sites in procuring a system suitable for their needs. Parameters assessed included SNR, geometric distortion, slice width and imaging speed.

Due to a change in policy within the NHS's Centre for Evidence-Based Purchasing (CEP) which moved away from technical reporting and evaluations, MagNET ceased operation in the late 2000s. The cultural legacy of MagNET as an independent assessment centre, however, is one that is still much discussed in the MRI community in the UK. As MRI evolves to encompass increasingly quantitative techniques, it is clear that the demand for a MagNET-like service will only increase.

The need for metrological rigour means that NMIs are increasingly becoming interested in supporting work in quantitative MRI. Traceable quantitative MRI needs involvement of both the research and clinical communities, including MR physicists, computer scientists, radiologists, radiographers, and patient groups. There is a multiplicity of voices within the field of MRI, and developing effective procedures require input from them all, placing everything in its clinical context. An appreciation of the needs of the user when designing experimental hardware and procedures allows for metrological rigour whilst avoiding overengineering. This need for improved metrology means that NMIs are increasingly becoming interested in supporting work in quantitative MRI, in addition to the efforts being made by larger consortia such as QIBA and EIBALL.

CONCLUSIONS

Quantitative MRI techniques are becoming increasingly important in the diagnosis, aetiology and treatment monitoring of many pathologies. In order to safely and confidently deploy them into regular use, however, there needs to be a similar level of trust in MR-derived biomarkers as there is in other biomarkers such as in the measurement of circulating tumour cells.

It is vital that the metrology of MR-derived biomarkers be implemented in order to have confidence in the veracity, not just in the form of experimental rigour and traceability measurements, but also in educating users in quantitative metrology as it applies to qMRI. It is crucial for the development of qMRI as a science to have complete engagement between all involved; from scanner manufacturers, metrologists and standards bodies all the way through to medical physicists and health-care practitioners. Without metrological underpinning, qMRI is a collection of isolated, niche techniques. With them, it provides a framework of novel biomarkers capable of supporting a new generation of clinical techniques, AI approaches, and personalised care.

REFERENCES

1. Cercignani S, Dowell M;, Tofts N. *Quantitative MRI of the brain: principles of physical measurement*. 2nd ed. Boca Raton, Florida: CRC Press; 2018.
2. BIPM. Joint Committee for Guides in Metrology JCGM 200: International vocabulary of metrology – Basic and general concepts and associated terms (VIM). 2012. Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf.
3. Smith NAS, Sinden D, Thomas SA, Romanchikova M, Talbott JE, Adeogun M. Building confidence in digital health through metrology. *Br J Radiol* 2020; **93**: 20190574. doi: https://doi.org/10.1259/bjr.20190574
4. Plant AL, Becker CA, Hanisch RJ, Boisvert RF, Possolo AM, Elliott JT. How measurement science can improve confidence in research results. *PLoS Biol* 2018; **16**: e2004299. doi: https://doi.org/10.1371/journal.pbio.2004299
5. Waldman AD, Jackson A, Price SJ, Clark CA, Booth TC, Auer DP, et al. Quantitative imaging biomarkers in neuro-oncology. *Nat Rev Clin Oncol* 2009; **6**: 445–54. doi: https://doi.org/10.1038/nrclinonc.2009.92
6. Hockings Pet al. Chapter 2: MRI biomarkers. In: Seberlich N, Gulani V, Campbell-Washburn A, Sourbron S, Doneva M. I, Calamante F, eds.*Quantitative magnetic resonance imaging (volume 1)*. Cambridge, Massachusetts: Academic Press; 2020.
7. Atkinson AJ, Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69**: 89–95. doi: https://doi.org/10.1067/mcp.2001.113989
8. Sirolli V, Pieroni L, Di Liberato L, Urbani A, Bonomini M. Urinary peptidomic biomarkers in kidney diseases. *Int J Mol Sci* 2020; **21**: 96. doi: https://doi.org/10.3390/ijms21010096
9. Ganepola GA, Rutledge JR, Suman P, Yiengpruksawan A, Chang DH. Novel blood-based microRNA biomarker panel for early diagnosis of pancreatic cancer. *World J Gastrointest Oncol* 2014; **6**: 22–33. doi: https://doi.org/10.4251/wjgo.v6.i1.22
10. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000; **92**: 205–16. doi: https://doi.org/10.1093/jnci/92.3.205
11. Marrugo-Ramírez J, Mir M, Samitier J. Blood-Based cancer biomarkers in liquid biopsy: a promising non-invasive alternative to tissue biopsy. *Int J Mol Sci* 2018; **19**: 2877. doi: https://doi.org/10.3390/ijms19102877
12. Pallante P, Pisapia P, Bellevicine C, Malapelle U, Troncone G. Circulating tumour cells in predictive molecular pathology: focus on drug-sensitive assays and 3D culture. *Acta Cytol* 2019; **63**: 171–81. doi: https://doi.org/10.1159/000496213
13. Marcuello M, Vymetalkova V, Neves RPL, Duran-Sanchon S, Vedeld HM, Tham E, et al. Circulating biomarkers for early detection and clinical management of colorectal cancer. *Mol Aspects Med* 2019; **69**: 107–22. doi: https://doi.org/10.1016/j.mam.2019.06.002
14. Winter JD, Estrada M, Cheng H-LM, Margaret Cheng HL. Normal tissue quantitative T1 and T2* MRI relaxation time responses to hypercapnic and hyperoxic gases. *Acad Radiol* 2011; **18**: 1159–67. doi: https://doi.org/10.1016/j.acra.2011.04.016
15. Warach S, Gaa J, Siewert B, Wielopolski P, Edelman RR. Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging. *Ann Neurol* 1995; **37**: 231–41. doi: https://doi.org/10.1002/ana.410370214
16. Burakiewicz J, Sinclair CDJ, Fischer D, Walter GA, Kan HE, Hollingsworth KG. Quantifying fat replacement of muscle by quantitative MRI in muscular dystrophy. *J Neurol* 2017; **264**: 2053–67. doi: https://doi.org/10.1007/s00415-017-8547-3
17. Gandon Y, Olivié D, Guyader D, Aubé C, Oberti F, Sebille V, et al. Non-Invasive assessment of hepatic iron stores by MRI. *Lancet* 2004; **363**: 357–62. doi: https://doi.org/10.1016/S0140-6736(04)15436-6
18. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. the cardiac arrhythmia suppression trial. *N Engl J Med* 1991; **324**: 781–8. doi: https://doi.org/10.1056/NEJM199103213241201
19. RSNA. Quantitative Imaging Biomarkers Alliance. 2020. Available from: https://www.rsna.org/en/research/quantitative-imaging-biomarkers-alliance.
20. ESR. ESR research committee. 2020. Available from: https://www.myesr.org/research/esr-research-committee#paragraph_grid_5924.
21. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* 2018; **166**: 400–24. doi: https://doi.org/10.1016/j.neuroimage.2017.10.034
22. Seraphim A, Knott KD, Augusto J, Bhuva AN, Manisty C, Moon JC. Quantitative cardiac MRI. *J Magn Reson Imaging* 2020; **51**: 693–711. doi: https://doi.org/10.1002/jmri.26789
23. Deoni SCL. Quantitative relaxometry of the brain. *Top Magn Reson Imaging* 2010; **21**: 101–13. doi: https://doi.org/10.1097/RMR.0b013e31821e56d8
24. Lescher S, Jurcoane A, Veit A, Bähr O, Deichmann R, Hattingen E. Quantitative T1 and T2 mapping in recurrent glioblastomas under bevacizumab: earlier detection of tumor progression compared to conventional MRI. *Neuroradiology* 2015; **57**: 11–20. doi: https://doi.org/10.1007/s00234-014-1445-9
25. Hattingen E, Müller A, Jurcoane A, Mädler B, Ditter P, Schild H, et al. Value of quantitative magnetic resonance imaging T1-relaxometry in predicting contrast-enhancement in glioblastoma patients. *Oncotarget* 2017; **8**: 53542–51. doi: https://doi.org/10.18632/oncotarget.18612
26. Lagergren K, Wahlin K, Mattsson F, Alderson D, Lagergren J. Haemochromatosis and gastrointestinal cancer. *Int J Cancer* 2016; **139**: 1740–3. doi: https://doi.org/10.1002/ijc.30229
27. Limdi JK, Crampton JR. Hereditary haemochromatosis. *QJM* 2004; **97**: 315–24. doi: https://doi.org/10.1093/qjmed/hch065
28. Litwack G. Chapter 19: Micronutrients (Metals and Iodine). In: *Human Biochemistry*; 2018. pp. 591–643. doi: https://doi.org/10.1016/C2009-0-63992-1
29. Hernando D, Levin YS, Sirlin CB, Reeder SB. Quantification of liver iron with MRI: state of the art and remaining challenges. *J Magn Reson Imaging* 2014; **40**: 1003–21. doi: https://doi.org/10.1002/jmri.24584
30. Labranche R, Gilbert G, Cerny M, Vu K-N, Soulières D, Olivié D, et al. Liver iron quantification with MR imaging: a primer for radiologists. *Radiographics* 2018; **38**: 392–412. doi: https://doi.org/10.1148/rg.2018170079
31. De novo classification request for ferriscan r2-mri analysis system decision summary. US Food and Drug Administration. Available from: https://www.accessdata.fda.gov/cdrh_docs/reviews/K124065.pdf.
32. Resonance Health. Ferriscan® – MRI measurement of liver iron concentration. 2018. Available from: https://www.resonancehealth.com/products/ferriscan-mri-measurement-of-liver-iron-concentration.html.

33. Henninger B, Alustiza J, Garbowski M, Gandon Y. Practical guide to quantification of hepatic iron with MRI. *Eur Radiol* 2020; **30**: 383–93. doi: https://doi.org/10.1007/s00330-019-06380-9
34. Wood JC, Enriquez C, Ghugre N, Tyzka JM, Carson S, Nelson MD, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood* 2005; **106**: 1460–5. doi: https://doi.org/10.1182/blood-2004-10-3982
35. Garbowski MW, Carpenter J-P, Smith G, Roughton M, Alam MH, He T, et al. Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. *J Cardiovasc Magn Reson* 2014; **16**: 40. doi: https://doi.org/10.1186/1532-429X-16-40
36. Noureddin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le T-A, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. *Hepatology* 2013; **58**: 1930–40. doi: https://doi.org/10.1002/hep.26455
37. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016; **64**: 73–84. doi: https://doi.org/10.1002/hep.28431
38. Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am* 2010; **18**: 337–57. doi: https://doi.org/10.1016/j.mric.2010.08.013
39. Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med* 2001; **344**: 495–500. doi: https://doi.org/10.1056/NEJM200102153440706
40. Dixon WT. Simple proton spectroscopic imaging. *Radiology* 1984; **153**: 189–94. doi: https://doi.org/10.1148/radiology.153.1.6089263
41. Brandão S, Seixas D, Ayres-Basto M, Castro S, Neto J, Martins C, et al. Comparing T1-weighted and T2-weighted three-point Dixon technique with conventional T1-weighted fat-saturation and short-tau inversion recovery (STIR) techniques for the study of the lumbar spine in a short-bore MRI machine. *Clin Radiol* 2013; **68**: e617–23. doi: https://doi.org/10.1016/j.crad.2013.06.004
42. Ma X, Holalkere N-S, Kambadakone R A, Mino-Kenudson M, Hahn PF, Sahani DV. Imaging-based quantification of hepatic fat: methods and clinical applications. *Radiographics* 2009; **29**: 1253–77. doi: https://doi.org/10.1148/rg.295085186
43. Chalela JA, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, et al. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet* 2007; **369**: 293–8. doi: https://doi.org/10.1016/S0140-6736(07)60151-2
44. Leinonen V, Vanninen R, Rauramaa T. Raised intracranial pressure and brain edema. *Handb Clin Neurol* 2018; **145**: 25–37. doi: https://doi.org/10.1016/B978-0-12-802395-2.00004-3
45. Axer H, Grässel D, Brämer D, Fitzek S, Kaiser WA, Witte OW, et al. Time course of diffusion imaging in acute brainstem infarcts. *J Magn Reson Imaging* 2007; **26**: 905–12. doi: https://doi.org/10.1002/jmri.21088
46. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. *Radiology* 1996; **201**: 637–48. doi: https://doi.org/10.1148/radiology.201.3.8939209
47. Tournier S. Chapter 8 - Moving Beyond DTI: High Angular Resolution Diffusion Imaging (HARDI). In: Mori S, Tournier S, eds.*Introduction to Diffusion Tensor Imaging*. 2. **vol. 9780123983985**: Academic Press; 2014.
48. Liu T, Li H, Wong K, Tarokh A, Guo L, Wong STC. Brain tissue segmentation based on DTI data. *Neuroimage* 2007; **38**: 114–23. doi: https://doi.org/10.1016/j.neuroimage.2007.07.002
49. Panesar SS, Abhinav K, Yeh F-C, Jacquesson T, Collins M, Fernandez-Miranda J. Tractography for surgical neuro-oncology planning: towards a gold standard. *Neurotherapeutics* 2019; **16**: 36–51. doi: https://doi.org/10.1007/s13311-018-00697-x
50. Zhou XJ, Leeds NE, Kumar AJ, Chong J, Levin VA. Differentiation of tumor recurrence from treatment-induced necrosis using quantitative diffusion MRI. *Proc Intl Soc Mag Reson Med* 2001; **9**: 726.
51. Costabile JD, Alaswad E, D'Souza S, Thompson JA, Ormond DR. Current applications of diffusion tensor imaging and tractography in intracranial tumor resection. *Front Oncol* 2019; **9**: 426. doi: https://doi.org/10.3389/fonc.2019.00426
52. Kraft KA, Fatouros PP, Clarke GD, Kishore PR. An MRI phantom material for quantitative relaxometry. *Magn Reson Med* 1987; **5**: 555–62. doi: https://doi.org/10.1002/mrm.1910050606
53. Keenan KE, Ainslie M, Barker AJ, Boss MA, Cecil KM, Charles C, et al. Quantitative magnetic resonance imaging phantoms: a review and the need for a system phantom. *Magn Reson Med* 2018; **79**: 48–61. doi: https://doi.org/10.1002/mrm.26982
54. CIRS. Large field MRI distortion phantom. 2020. Available from: https://www.cirsinc.com/products/radiation-therapy/large-field-mri-distortion-phantom/.
55. Gunter JL, Bernstein MA, Borowski BJ, Ward CP, Britson PJ, Felmlee JP, et al. Measurement of MRI scanner performance with the ADNI phantom. *Med Phys* 2009; **36**: 2193–205. doi: https://doi.org/10.1118/1.3116776
56. Keenan K, Stupic KF, Boss MA, Russek SE. Comparison of T1 measurement using ISMRM/NIST system phantom in Proceedings of the 24th Annual Meeting of ISMRM. p. 3290. 2016. Available from: https://www.nist.gov/publications/comparison-t1-measurement-using-ismrmnist-system-phantom.
57. gov.uk. Magnetic resonance imaging equipment in clinical use: safety guidelines. 2014. Available from: https://www.gov.uk/government/publications/safety-guidelines-for-magnetic-resonance-imaging-equipment-in-clinical-use.
58. IPEM. POLICY STATEMENT: The role of the Clinical Scientist in Magnetic Resonance Imaging units conducting human diagnostic imaging. 2017. Available from: https://www.ipem.ac.uk/ScientificJournalsPublications/IPEMStatementsandNotices.aspx.
59. De Wilde J, Price D, Curran J, Williams J, Kitney R. Standardization of performance evaluation in MRI: 13 Years' experience of intersystem comparison. *Concepts Magn Reson* 2002; **15**: 111–6. doi: https://doi.org/10.1002/cmr.10012