6-8 (2023) 100019

Contents lists available at ScienceDirect

IPEM-Translation

journal homepage: www.elsevier.com/locate/ipemt

Full Length Article

The perfect diagnostic imaging machine and what it means for quantitative MRI reproducibility☆

Matt G. Hall [a,b,*], Matthew T.D. Cashmore [a], Cormac McGrath [c], Aaron McCann [c], Paul S. Tofts [d]

[a] National Physical Laboratory, Teddington, UK
[b] GOS Institute of Child Health, University College London, UK
[c] Northern Ireland Regional Medical Physics Service, Belfast Health and Social Care Trust, UK
[d] Brighton and Sussex Medical School, UK

ARTICLE INFO

Keywords:
MRI
Quantitative MRI
Standardisation
Harmonisation
Perfect machine
Reproducibility
Accreditation
Site qualification

ABSTRACT

Starting from the recently published idea of the "perfect machine", we argue that quantitative MRI (qMRI) supported by a rigorous metrological framework could not only drastically improve reproducibility in MRI and support large-scale studies, there is also scope to accredit individual MRI scanners in particular applications via a suitable accreditation scheme. We present the idea of the perfect diagnostic imaging machine, including describing the background of the ideas and how they lead to the idea of accreditation in qMRI. The scheme presented here is not intended to be the last word in accreditation, but to stimulate debate in the idea and whether or not is has merit for qMRI and its clinical and research context.

Introduction

Magnetic Resonance Imaging (MRI) is an advanced 3D medical imaging modality in widespread use in clinical care and larger scale studies. It is highly versatile, with choices of pulse sequence providing different image contrast mechanisms and an unrivalled ability to differentiate soft tissues [1]. MRI can provide measures of dynamical processes such as diffusion [2], and assess the chemical makeup of tissue [3,4].

Typically, MRI contrast as used in the clinic is not quantitative. Images are formed by "weighting" voxel intensities via some chosen mechanism rather than providing estimates of physical parameters [1]. This is fast and effective but can be challenging for reproducibility. Conventional MR images from different scanners—even after great care has been taken to harmonise the protocols—are rarely directly comparable, and this has led to a culture of single-use images, used once by human experts, and then archived or discarded. It also requires a more operator intervention to achieve optimal image quality than other medical imaging methods.

Single use images are problematic when assessing treatment efficacy [5,6], considering long-term conditions such as dementia [7], or for larger studies such as clinical trials or population-level research projects [8]. Considerable time and effort are spent harmonising imaging protocols, and imaging referrals are complicated by the need to scan a patient on the same hardware as previous scans. Even when many scanners are successfully harmonised, diffuse changes in signal are challenging to detect, and the biological significance of an observed change in an image is not always clear.

Enter quantitative MRI (qMRI). qMRI forms images which contain measurements of a physical or physiological parameter. Often this is a map of the value of a parameter which can be measured from an appropriate set of MRI data. T1 and T2 are good examples, but it could just as easily be of fat fraction, iron content, apparent diffusion coefficient, or many others. In each case we are using the scanner to make measurements, rather than just obtain relative contrast [4,9].

Measurements are fundamentally different from conventional images because the parameters of interest have an existence independent of the image formation process. This means they can be measured in more than one way, potentially including both MRI- and non-MRI-based approaches. This allows practical, front line imaging methods to be compared to reference measurements made under highly controlled conditions, potentially without imaging at all. This is an important

☆ The work described in this paper is based on a presentation delivered by the author at IPEM's inaugural Science, Technology and Engineering Forum (STEF), Glasgow, UK, 28th February–1st March 2023.

* Corresponding author.
E-mail address: matt.hall@npl.co.uk (M.G. Hall).

https://doi.org/10.1016/j.ipemt.2023.100019
Received 14 June 2023; Received in revised form 6 November 2023; Accepted 7 December 2023
Available online 20 December 2023

difference, because it means qMRI has far greater potential for reproducibility.

Quantitative Imaging leads naturally to the concept of a biomarker, i.e. ``an objective characteristic derived from an in vivo image measured on a ratio or interval scale as an indicator of a normal biological process, a pathogenic process, or a response to a therapeutic intervention.'' [10a]. In the context of qMRI, this means agreement on the choice and definition of measurand. Examples here might be T1 and T2, which are values which depend on the main field strength and temperature. Other choices such as fat or iron content are potentially verifiable via non-MRI methods. A key aspect of this is that each biomarker have an agreed, documented definition—T1 and T2 for example need definitions not just in terms of relaxation dynamics, but also which pulse sequences are used to measure them and what model used to extract them. This is the role of guidance and formal standards.

Quantification allows a particular imaging process and analysis pipeline to be assessed by considering how well it performs and means that the performance of two different measurement pipelines can be meaningfully and quantitatively compared. This article will argue that by using this as a base, we can go further and quantify effect size and sensitivity across different scanners and compare them with effect sizes measured from lab-based methods to assess performance of MRI hardware in applications to specific biomedical questions.

The strengths of quantitative imaging do not come for free, however. To ensure consistency, measurement equipment and processes need to be traceable to internationally-agreed definitions of the Système international (SI) units (the metre, second, kilogram, kelvin, mole, ampere, and candela) from which all other units are derived. A measurement can be said to be "traceable" if there is an unbroken chain of calibrations between the measuring equipment and procedure being used and the relevant SI unit. This is performed by calibrating to local measuring equipment and secondary standards, which are in turn calibrated to primary standards which are designed to most accurately realise the SI or derived units they represent [10,11]. In principle there can be more than one primary standard for any given measurand, for example in different countries.

This also connects naturally to the idea of an "uncertainty", which is the limits of accuracy of a given measurement. Uncertainties are which are conceptually distinct from errors since they do not require knowledge of a hypothetical ideal value. They can be quantified by understanding all aspects of the measurement process in detail. A quantified uncertainty on two different measurements allows them to be compared even if the measurement processes differ. Knowledge of uncertainty means that a practical and affordable measurement can be quantitatively compared to a more accurate but less practical method. This allows traceability.

This kind of metrological framework is common in many areas of science, engineering, and industry. SI Traceability is the reason we can trust a 10 mm spanner to fit a 10 mm nut regardless of when or where it was manufactured. It is also the reason why a 5 mg dose of a pharmaceutical is always the same regardless of which factory produced it. In healthcare applications, MRI is comparatively unusual in not conforming to this kind of framework and we need only look as far as Radiotherapy to see a similar approach, with both treatment planning and radiation dose both being subject to routine audit and calibration [12–14].

Providing SI traceability for MRI requires a chain of reference measurements ultimately back to a primary standard for each different kind of MRI measurement. For the examples we have given so far—T1, T2, chemical composition and other measurements which can be described by in a single image voxel—this can be provided in a similar way to the Caliber-NIST system phantom [15]. Here a phantom is provided to sites along with a recommended imaging protocol and analysis software to extract parameter values. The contrast-generating materials in the phantoms provide a well-characterised reference for a qMRI parameter of interest (T2, say) because they have been characterised on NIST's SI-traceable variable field NMR spectrometer which has field strength, pulse timings, and sample temperature calibrated to the US nation primary standards for frequency, time, and temperature. This enables measurements from individual scanners to compared and benchmarked against this metrologically robust reference. As of 2023, NPL in the UK is also building a similar facility to provide reference metrology for MRI.

Image-based measures such as the size, shape, or position or tissue features are another important class of measurand. Because these measurements are spatially extended, reference metrology must cannot be performed on a spectrometer (which cannot form images) and instead dimensional metrology must be used. MR-guided radiotherapy is a key example. MRI's soft tissue contrast makes it an attractive proposition for Radiotherapy treatment planning, but this highlights a need for improved quantification in MRI [16]. Spatial distortion is unavoidably present in any MR image, partly due to degradation of static field homogeneity and gradient linearity as we move from the isocentre and partly because the field and gradients are distorted by susceptibility differences in the tissue of the patient being imaged.

Distortion is unavoidable in any MRI application, but in Radiotherapy treatment planning [17,16] it represents a source of bias and uncertainty in the positions and sizes of all tissue features, including tumours and organs at risk and means having to include wider margins in treatment planning [18]. This deeply problematic and can lead to a reduced dose to the primary tumour volume, and potentially irradiate healthy surrounding tissue [13].

The effect of distortion and the performance of distortion correction algorithms employed need to be quantified and included in any uncertainty budgets derived from facility audits. Currently this is only partially possible, and hence challenging when characterising the uncertainty of dose delivery on an MR linac. Addressing this requires new test objects, procedures, and approaches.

Reference metrology can also provide important information about reproducibility and inter-subject variability. Quantitative contrast, appropriately characterised and benchmarked, leads to the ability to compare two different measurements and quantify how significant an observed difference is. This relates directly to the design of clinical trials—for a given effect size we can now quantify how many trial participants would be required to observe an effect. Well-characterised quantitative measurements reduce inter-scanner variability and hence increase statistical power. This drives the required number of participants down, reducing cost and duration without sacrificing rigour.

In the context of MRI, the word "calibration" is problematic. MRI scanners are highly versatile machines capable of forming images in many different ways with a broad range of contrasts. Even a conventional MR image without quantitative contrast can potentially be used in a quantitative way, such as measuring the volume or position of a feature. As such, there is no single procedure which can calibrate a scanner. Instead, each form (and potentially each application) of qMRI needs its own benchmarking and "calibration". Throughout this article, we will avoid the term calibration when talking about solely MR-based measurements and instead refer to characterisation, and specifically with this application-specific meaning in mind.

The ability to quantify variability due to a single scanner, and the related capability to assess variability between scanners (each in an individual application) means it becomes possible to compare measurement variability to the inherent variability in a patient population. This is at the heart of the concept of the "perfect machine" as detailed by Tofts [19]. It allows us to define quantitative criteria for the performance of qMRI measurements for different applications.

Quantitative MRI and the need for metrology

Metrology is the science and practice of measurement. It involves the study, development, and application of methods, techniques, and instruments used for precise and accurate measurement of physical quantities. Metrology plays a crucial role in various fields, including

manufacturing, engineering, science, and technology. The goal of metrology is to ensure that measurements are reliable, consistent, and traceable to internationally recognized standards. It encompasses the entire measurement process, including the design of measurement systems, characterisation of measuring instruments, and the establishment of measurement standards and units.

Any measurement contains three important components: the quantity of the parameter under investigation, a unit, and an indication of the associated measurement uncertainty, including an expression of confidence. Uncertainties represent the limits of knowledge of a quantity provided by a measurement. Since it is impossible to obtain the "true" value of the parameter, the uncertainty expresses the range around the stated measurement in which the true value is expected to lie, with the confidence level giving the statistical confidence of that determination. With quantified uncertainties, two measurements of the same quantity resulting from different measurement processes can be meaningfully compared. Without uncertainties the significance (or otherwise) or the difference between two measurements cannot be assessed.

Consistency of units is achieved via the use of standards. This term is sometimes problematic because it can refer to several different things, but for current purposes a standard in metrology is the realisation of the definition of a unit of measurement. A *primary* standard is one that is not calibrated subject to any other standards. Historically, these have often been physical artifacts, such as the International Prototype metre. Since 2018, however, SI units are defined in terms of fundamental physical constants [20].

The concept of the perfect diagnostic imaging machine

The perfect machine is a statistical concept which defines criteria for measurement performance beyond which uncertainty no longer needs to be improved. Once the uncertainty associated with a measurement process is reduced below the natural variability of a measurand in a target population, the measuring equipment no longer significantly impacts a set of measurements and the measuring device can be considered "perfected". This criterion is clearly dependent on the application of the measurement and the variability in the target population. A "perfect" machine in one application may not be perfect in another.

The concept allows us to consider diagnostic imaging—the "perfect quantitative diagnostic imaging (qDI) machine". Here we measure biomarkers from images to inform decisions about the presence or progress of various diseases. We could differentiate between perfect MRI machines, perfect CT machines, etc. whilst still retaining the central concept. In each case we acquire images of a population in which a particular disease may or may not be present. The disease may be characterised by a tissue feature (such as a tumour or lesion), or a change in the distribution of some parameter of interest. Genuinely quantitative and reproducible parameter maps also allow the definition of normative ranges, allowing the detection of more diffuse changes to tissue in otherwise-normal appearing images. A tissue may appear normal on an image, yet careful measurement of a qMR parameter may show that it has a value outside of the normal range [21,22].

Intrinsic variability

Where a disease is present, it will manifest differently in different individuals. Tumours fall into classes but no two are ever exactly alike. A change in fat content in the liver may allow differentiation between treatments or disease types but the exact level will vary from individual to individual. Since these are physically meaningful characteristics, they can in principle be measured not just through imaging but also via other methods such as biopsy. In some cases, this allows the level of variability in patients to be measured and quantified as part of research into the disease. This is the intrinsic variability of the disease and although it may be a complex or difficult to measure, it only needs to be done once.

Measurement uncertainty

Metrology conventionally recognises two major classes of uncertainty; Type A and Type B. Type A uncertainties are those which arise through statistical sampling of data and what is commonly thought of as "noise". A Type B uncertainty is one which is not dependent on a statistical analysis of data, and often is associated with non-local factors to the measurement process such as performance drift over time.

The uncertainty in any measurement process can be compared to the natural biological variability in a patient population. Variability between patients is not a function of a measurement process, it is an aspect of the measurand itself. The amount of variability present in patient populations depends on the disease, the inherent variability in human physiology, and the type of study being performed. Comparisons between groups involve a different level of variability that comparisons between individuals, for example. This leads naturally to different levels of certification for different types of application.

By comparing measurement uncertainty to intrinsic variability we can define the perfect machine. Quantitative diagnostic imaging allows us to define a measurement uncertainty on a medical imaging process, and hence for the images to be compared to the variability in the patient population. This allows us to assess the statistical significance of a change in a patient's images and assess whether imaging performance is sufficient for a particular application.

Quantification and accreditation

Tofts [19] defines four levels of performance requirements: platinum, gold, silver, and bronze (in descending order of stringency on the uncertainty requirement). These are defined by comparing the uncertainty to the inherent variability. This makes it possible to benchmark measurement performance in a given application, place the facility in a given category, and further to some form of certification on this basis. This would have to be done periodically to account for hardware degradation and scanner updates, but in principle sites could seek accreditation in particular applications to demonstrate compliance with a certain standard of measurement. A given site might be accredited at different levels for different applications or seek to implement new procedures to gain accreditation in a new one.

The most obvious benefit of an accreditation scheme is that it becomes easier to select sites for studies. Currently any large-scale study must spend an extended period of time implementing and harmonising image acquisition and processing across multiple scanners on multiple sites. Accrediting sites in individual imaging applications means that images from similarly accredited sites can be quantitatively compared and are already harmonised via the accreditation scheme. This immediately simplifies trial qualification and reduces the amount of work required to set up individual studies. Quantified measurement uncertainties are also helpful for study design. The uncertainty of each accreditation band can be used to assess study power and provide an accurate estimate of the number of sites and patients required.

Accreditation of imaging in specific applications also has benefits clinically. Quantitative imaging and the associated comparability provides assurance that scans taken on similarly accredited facilities are directly comparable, and as such that facilities are equivalent - a patient does not necessarily have to return to the same scanner for a follow-up scan. This simplifies scheduling and potentially means shorter waiting times for appointments. Moreover, even single point qMRI measurements have the capacity to alter treatment pathways—although metrological support is essential.

An established methodology for benchmarking scanner performance is also valuable to the planning and development of new facilities. As scanner characterisation is performed per application, it provides valuable information on the varying strengths individual scanner models may demonstrate. A radiotherapy treatment planning facility may have little need for gold-tier diffusion status and can prioritise

systems which have a demonstrable proficiency in repeatably accurate spatial extent measurements. This allows for centres to understand what the most suitable options for purchase and development is, and the level of performance they can expect to deliver.

Accreditation also has potential benefits for equipment procurement and scanner commissioning. By providing quantitatively verifiable standards it allows hospitals to identify high priority applications for their patients and identify scanner makes and models which are known to be accredited elsewhere. This is also potentially beneficial to manufacturers as it provides common standards to demonstrate performance against, rather than having to demonstrate performance to separate specifications from each individual customer.

Gap analysis

There is currently very little uniformity in the implementation of assessment and quality assurance (QA) of MRI scanners for quantitative techniques. One of the most commonplace of these is in the use of the ACR large phantom [23] and associated guidance for the characterisation of basic image parameters such as spatial resolution and slice thickness. The most significant drawbacks to this are that most procedures involved require subjective judgement by the user (and therefore is not fully quantitative) and that it only covers spatial distortion. To robustly assess measurement uncertainty, quantitative references are needed, and this means traceable phantoms.

There is a wide range of phantoms available both commercially and in research literature, however the majority of these do not come with associated metrology and longitudinal stability measurements. This can be challenging due to the competing needs of the underlying measurement science and clinical implementation. Traceable MRI phantoms are increasingly available however, notable examples being the Caliber-ISMRM-NIST phantom [15] and the KRISS modular MRI phantom [24].

The ideal scenario for qMRI would be one in which it is straightforward to compare multiple measurements from different times and locations. This could manifest in being able to understand how diffusion measurements made on a 3T Siemens scanner in London at one temperature, and one from a 1.5T Philips machine in Belfast at another, both relate not only to each other, but the SI definitions of the metre and second. This presents a huge challenge to both metrologists and clinical scientists. The first step is to traceably link the measurements of interest to their units. This requires a traceability chain for each measurand of interest: relaxometry requires a path to the SI unit of time, for tissue composition measurements we need a route to the mole or kilogram. Primary standard metrology for these base units are well-established, and the mainstay of National Metrology Institutes such as NPL (National Physical Laboratory, UK), NIST (National Institute of Standards and Technology, US) and PTB (Physikalisch-Technische Bundesanstalt, Germany). It also requires consensus in the community on the definitions of the parameters being measured—T1, for example, is currently defined as the time constant associated with a relaxation process, but there would also need to be consensus over how best to measure it. Does an inversion recovery performed on a carefully calibrated spectrometer represent a gold standard measurement, and what procedure should be used to estimate the parameter from the data? The same is the case for other MRI measurands. Thus we can foresee a set of spectrometers at different locations, all making primary standard measurements according to an internationally-agreed procedure.

Beyond the MRI-related aspects of the measurements, it is also important to know how phantoms change over time. A gel phantom may degrade, paramagnetic ions in relaxometry samples may be leached into the container, a length scale may absorb water and warp. Phantoms can be costly and may only be viable for a relatively short lifespan. Primary standard traceability measurements on every phantom produced is not financially viable. It is possible however to improve the accessibility of well characterised test objects through a calibration chain. It is not necessary for every centre to own a rigorously understood test object and can rely on periodic calibrations against a small number of well-maintained reference objects hosted by an external institute.

Metrological tools to support phantom development are in their infancy and qMRI techniques are complicated. A rigorous understanding of measurement uncertainty means considering every source of variation which may affect the results. This is fiendishly difficult for MRI.

One of the complicating factors is the range of different recommendations available. There is very little international standardisation in qMRI. The European implementation of IEC 60601[1] specifically excludes MRI from the status of measurement device and in the US scanners ship with a notice explicitly stating they are not measurement devices. This means that current standards are silent on quantitative methods in MRI. Given the groundswell of interest in qMRI, this is a precarious position. In the absence of high-level clarity and standardisation there is variation in the adoption of guidance made in various consensus standards. For example, whilst the ACR accreditation procedure guidelines are commonly used in the acceptance testing and commissioning of scanners, this is not always the case. MHRA guidance is limited to recommending that QA and commissioning take place, with no comment on what that should involve. This means that scanners throughout the country are held to their own unique benchmark, and to compare QA results requires in depth understanding on the details of how QA is performed at each site.

Recommendations

We believe that the case for some form of assessment of measurement performance for quantitative MRI is compelling and worth considering as a community. We do not need to look far afield for a potential model, however – therapy and ionising radiation frameworks provide a useful example to follow. Metrology for radiotherapy is a well-developed field, and the model of primary standards, secondary standards, quantification and audit provides a useful model for developing metrology for qMRI, and potentially for other forms of in-vivo metrology supporting other emerging techniques.

Whether that extends to formal accreditation and badging for qMRI is a matter for debate but having established the principle of quantified measurement uncertainty defined in relation to an intrinsic variation in a patient population leads to the concept of accreditation quite naturally, and has benefits across imaging applications, particular in large scale studies such as clinical trials. Overall, any metrological or accreditation service also needs to be supported by a suitable cost-benefit analysis which shows how it will impact patient care. In many cases, current practices may well prove sufficient, but as we have highlighted there is significant potential for long-term conditions or for larger studies such as clinical trials where variability between scanners of timepoints becomes important.

Accreditation also raises the question of an accrediting body. This would ideally be an ISO-9000 accredited organisation, to be perceived as politically neutral. In the UK, NPL already acts in this role for Radiotherapy audit, but an alternative model could be a centralised NHS facility. A potential historical model for this would be MagNET [25], an MRI quality assurance service provided by and to the NHS from 1988 to 2009.

All of this requires a metrological framework to realise, whereby scanner-based measurements are benchmarked and referenced to national and international standards in very much the same way that other areas of imaging and measurement are currently benchmarked. Notice also that this scheme rests on specifying measurement performance, not on specifying the measurement process itself. If standards of uncertainty are agreed, different measurement pipelines can be meaningfully compared and assessed and there is no need to specify the exact method.

[1] https://www.bsigroup.com/en-GB/medical-devices/our-services/en-60601/

It therefore requires consensus on how to assess measurement performance and best practice guidance and training on how to perform it. Arguably, it also requires an accessible library of traceably characterised phantoms.

The MRI community is currently at a tipping point. Interest in standardisation is increasing and available metrological support is improving. Increased needs for reproducibility are presenting themselves in the form of AI and the need to improve the utilisation of scanner resources. Quantification is the key to achieving all of this, and to fully realise its potential, some form of accreditation scheme feels essential. We have presented one possible scheme with the hope that it will provoke debate and drive new ideas and uptake.

Ethical approval

Not required.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

Work on this manuscript was performed in connection with the iMet-MRI project. This 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.

References

[1] D. McRobbie, E. Moore, M. Graves, M. Prince, MRI from Picture to Proton, second ed., Cambridge University Press, Cambridge, 2006.
[2] H. Johansen-Berg, T.E.J. Behrens, Diffusion MRI: From Quantitative Measurement to In Vivo Neuroanatomy, Academic Press, 2014.
[3] T.J.P. Bray, M.D. Chouhan, S. Punwani, A. Bainbridge, M.A. Hall-Craggs, Fat fraction mapping using magnetic resonance imaging: insight into pathophysiology, Br. J. Radiol. 91 (1089) (2018) 20170344.
[4] M. Cercignani, N.G. Dowell, P.S. Tofts, Quantitative MRI of the Brain: Principles of Physical Measurement, second ed., CRC Press, 2018.
[5] F. Fazekas, P. Soelberg-Sorensen, G. Comi, M. Filippi, MRI to monitor treatment efficacy in multiple sclerosis, J. Neuroimaging 17 (suppl. 1) (2007) 50S–55S.
[6] Y. Zhang, J. Yu, The role of MRI in the diagnosis and treatment of gastric cancer, Diagn. Interv. Radiol. 26 (3) (2020) 176–182.
[7] R. Schmidt, D. Havas, S. Ropele, C. Enzinger, F. Fazekas, MRI in dementia, Magn. Reson. Imaging Clin. N. Am. 18 (1) (2010) 121–132.
[8] T.J. Littlejohns, J. Holliday, L.M. Gibson, S. Garratt, N. Oesingmann, F. Alfaro-Almagro, J.D. Bell, C. Boultwood, e. al, The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions, Nat. Commun. 11 (1) (2020) 2624.
[9] N. Seiberlich, V. Gulani, A. Campbell-Washburn, S. Sourbron, M. Doneva, F. Calamante, H.H. Hu, Quantitative Magnetic Resonance Imaging, Elsevier, 2020.
[10] W. Gao, Metrology, Springer-Nature, 2019.
[11] BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML, International vocabulary of metrology—Basic and general concepts and associated terms (VIM), in: Joint Committee for Guides in Metrology, JCGM 200, 2012.
[12] C.H. Clark, N. Jornet, L.P. Muren, The role of dosimetry audit in achieving high quality radiotherapy, Phys. in Imaging Radioat. Oncol. 21 (5) (2018) 85–87.
[13] B.J. McParland, Medical Radiation Dosimetry, Springer-Nature, 2014.
[14] D. Eaton, G. Bass, P. Booker, J. Byrne, S. Duane, J. Frame, M. Grattan, R. A. Thomas, N. Thorp, A. Nisbet, IPEM code of practice for high-energy photon therapy dosimetry based on the NPL absorbed dose calibration service, Phys. Med. Biol. 65 (19) (2020), 195006.
[15] K.F. Stupic, M. Ainslie, M.A. Boss, C. Charles, A.M. Dienstfrey, J.L. Evelhoch, P. Finn, Z. Gimbutas, J.L. Gunter, D.L.G. Hill, C.R. Jack, E.F. Jackson, T. Karaulanov, K.E. Keenan, et al., A standard system phantom for magnetic resonance imaging, Magn. Reson. Med. 86 (3) (2021) 1194–1211.
[16] R. Speight, M. Tyyger, M.A. Schmidt, G. Liney, R. Johnstone, C.L. Eccles, M. Dubec, B. George, A. Henry, T. Herbert, T. Nyholm, F. Mahmood, J. Korhonen, R. Sims, R. H.N. Tijssen, F. Vanhoutte, S. Busoni, T. Lacornerie, H. McCallum, IPEM topical report: an international IPEM survey of MRI use for external beam radiotherapy treatment planning, Phys. Med. Biol. 66 (7) (2021), 075007.
[17] A. Walker, G.P. Liney, P.E. Metcalfe, L. Holloway, MRI distortion: considerations for MRI based radiotherapy treatment planning, Aust. Phys. Eng. Sci. Med. 37 (1) (2014) 103–113.
[18] L. Lu, X. Yang, B. Raterman, X. Jiang, M. Meineke, J. Grecula, D. Blakaj, J. Palmer, R. Raval, E. Thomas, D. Hintenlang, N. Gupta, Assessment of MRI image distortion based on 6 consecutive years of annual QAs and measurements on 14 MRI scanners used for radiation therapy, J. Clin. Med. Phys. 24 (1) (2023) e13843.
[19] P.S. Tofts, The perfect qMR machine: measurement variance much less than biological variance, Phys. Medica 104 (2022) 145–148.
[20] BIPM - International Bureau of Weights and Measures, Le Système international d'unités [The International System of Units], ISBN 978-92-822-2272-0, 2019.
[21] G.R. Davies, D.J. Tozer, M. Cercignani, A. Ramani, C.M. Dalton, A.J. Thompson, G. J. Barker, P.S. Tofts, D.H. Miller, Estimation of the macromolecular proton fraction and bound pool T2 in multiple sclerosis, Mult. Scler. 10 (6) (2004) 607–613.
[22] G.F. Piredda, S. Caneschi, T. Hilbert, G. Bonanno, A. Joseph, K. Egger, J. Peter, S. Klöppel, E. Jehli, M. Grieder, J. Slotboom, D. Seiffge, M. Goeldlin, R.R. Hoepner, T. Willems, e. al, Submillimeter T1 atlas for subject-specific abnormality detection at 7T, in: Proceedings of the Meeting of the International Society for Magnetic Resonance in Medicine, Tprpmtp, Canada, 2023.
[23] ACR, *ACR. Phantom Test Guidance for Use of the Large MRI Phantom for the ACR MRI Accreditation Program,* 2018.
[24] H.-M. Cho, C. Hong, C. Lee, H. Ding, T. Kim, B. Ahn, LEGO-compatible modular mapping phantom for magnetic resonance imaging, Sci. Rep. 10 (2020) 14755.
[25] D. Price, "MagNET - MRI Evaluation, Quality Assurance, Specification and Safety (archived)," 2000. [Online]. Available: https://web.archive.org/web/20090419053559/http:/www.magnet-mri.org/index.htm.