Physica Medica 101 (2022) 165–182

Contents lists available at ScienceDirect

Physica Medica

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

Clinical translation of quantitative magnetic resonance imaging biomarkers – An overview and gap analysis of current practice

Penny L. Hubbard Cristinacce [a,*], Sam Keaveney [b,c], Eric O. Aboagye [d], Matt G. Hall [e], Ross A. Little [a], James P.B. O'Connor [a,c], Geoff J.M. Parker [f,g], John C. Waterton [g,h], Jessica M. Winfield [b,c], Maite Jauregui-Osoro [d]

[a] Division of Cancer Sciences, The University of Manchester, Manchester M13 9PT, UK
[b] MRI Unit, Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT, UK
[c] Division of Radiotherapy and Imaging, The Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP, UK
[d] Department of Surgery & Cancer, Division of Cancer, Imperial College London, W12 0NN London, UK
[e] National Physical Laboratory, Hampton Road, Teddington TW11 0LW, UK
[f] Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, 90 High Holborn, London WC1V 6LJ, UK
[g] Bioxydyn Ltd, Manchester M15 6SZ, UK
[h] Division of Informatics, Imaging and Data Sciences, The University of Manchester, Manchester M13 9PT, UK

ARTICLE INFO

Keywords:
Quantitative MRI
QAQC
Clinical translation
Metrology

ABSTRACT

Purpose: This overview of the current landscape of quantitative magnetic resonance imaging biomarkers (qMR IBs) aims to support the standardisation of academic IBs to assist their translation to clinical practice.
Methods: We used three complementary approaches to investigate qMR IB use and quality management practices within the UK: 1) a literature search of qMR and quality management terms during 2011–2015 and 2016–2020; 2) a database search for clinical research studies using qMR IBs during 2016–2020; and 3) a survey to ascertain the current availability and quality management practices for clinical MRI scanners and associated equipment at research institutions across the UK.
Results: The analysis showed increased use of all qMR methods between the periods 2011–2015 and 2016–2020 and diffusion-tensor MRI and volumetry to be popular methods. However, the "translation ratio" of journal articles to clinical research studies was higher for qMR methods that have evidence of clinical translation via a commercial route, such as fat fraction and T_2 mapping.
The number of journal articles citing quality management terms doubled between the periods 2011–2015 and 2016–2020; although, its proportion relative to all journal articles only increased by 3.0%. The survey suggested that quality assurance (QA) and quality control (QC) of data acquisition procedures are under-reported in the literature and that QA/QC of acquired data/data analysis are under-developed and lack consistency between institutions.
Conclusions: We summarise current attempts to standardise and translate qMR IBs, and conclude by outlining the ideal quality management practices and providing a gap analysis between current practice and a metrological standard.

1. Introduction

Quantitative magnetic resonance (qMR) methods have increased in number, scope and popularity over the past two decades, with significant developments in both hardware and software improving the quality and speed of image acquisition and the physiological characteristics that can be extracted from MR data. At the same time, the potential of qMR measurements to improve understanding and diagnosis of disease has evolved, in tandem with the increased focus on personalised medicine. Relaxation times, diffusion, microstructure, perfusion, flow, chemical exchange, fat and iron content, susceptibility, temperature, metabolism, inflammation, fibrosis, elastic properties and chemical composition are some of the many properties that qMR is able to probe [1]. However, the theoretical potential of qMR has not translated into widespread clinical

* Corresponding author at: Division of Cancer Sciences, The University of Manchester, Manchester M13 9PT, UK.
E-mail address: penny.cristinacce@manchester.ac.uk (P.L. Hubbard Cristinacce).

https://doi.org/10.1016/j.ejmp.2022.08.015
Received 9 March 2022; Received in revised form 5 August 2022; Accepted 17 August 2022
Available online 30 August 2022
1120-1797/© 2022 Associazione Italiana di Fisica Medica e Sanitaria. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

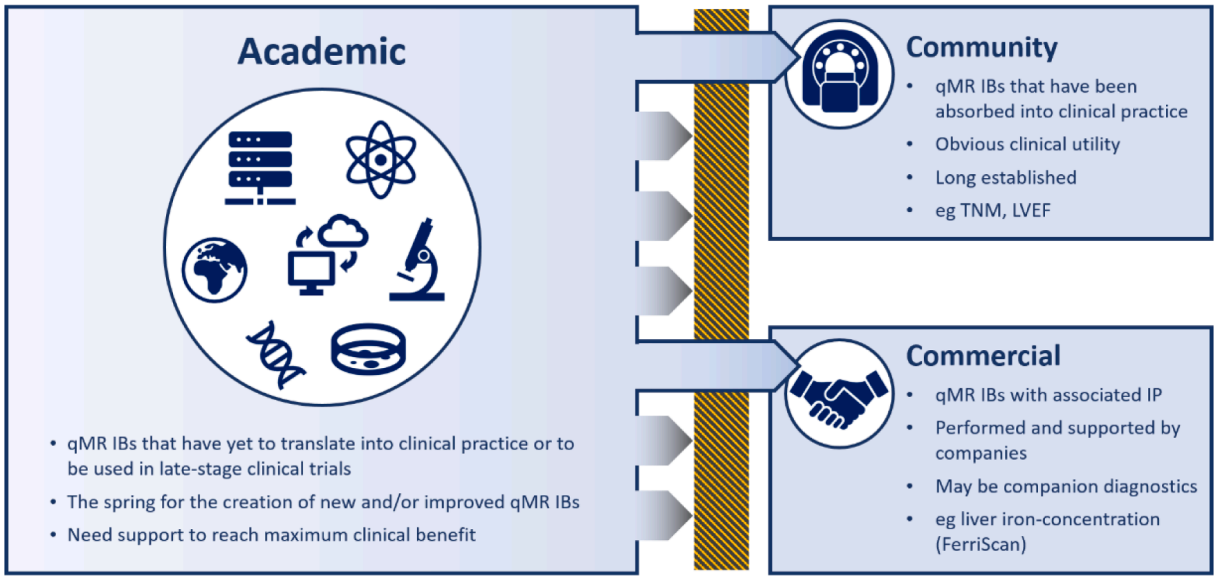


Fig. 1. Schematic showing the 3 main types of qMR IBs and their possible routes into clinical practice and late-stage clinical trials. Abbreviations: TNM (tumour: node: metastasis), LVEF (Left Ventricular Ejection Fraction), IP (Intellectual Property).

adoption, with a very limited number of qMR imaging/spectroscopic biomarkers (IBs) guiding clinical decisions [2,3].

The majority of qMR techniques require the acquisition of multiple images followed by fitting a model to the acquired data. This requires more time (and therefore cost) to acquire and analyse than conventional qualitative MRI, such as T_1 and T_2-weighted images, which maximise the contrast difference between the tissues of interest on a single image ready for expert radiological interpretation. Typical standard-of-care images can be obtained using different makes of scanner with different acquisition parameters. This variability in image acquisition is often tolerable provided the radiologist can arrive at the correct diagnosis by inspecting a picture. However, such divergence in acquisition or analysis – which arises from differences in scanner vendor, make/model, hardware performance, software version, field strength and age – is not tolerable in qMR as it may alter the numerical values of the qMR IB and will compromise the outcome of multicentre studies using qMR IBs as endpoints in clinical trials, [4] or translational studies to establish qMR IB cut-offs for use in clinical decision-making.

qMR IBs have multiple potential uses. They may enable earlier diagnosis and prognosis, often complementing or replacing biopsy [5]. They can provide measurements of beneficial or harmful response to treatment [6] and in some cases are acceptable to regulatory authorities as surrogate endpoints. "Predictive" qMR IBs support personalised medicine by indicating the most beneficial treatment for the individual patient, monitoring, treatment planning, and by helping to ensure that patients avoid treatments where harm outweighs the benefits [7]. To complement conventional molecular "biospecimen" biomarkers, qMR typically uses images with physically and physiologically meaningful metrics that quantitatively describe characteristics such as microstructure, perfusion, metabolism, function, inflammation or fibrosis. However, such quantification requires standardisation, consistent acquisition and analysis and rigorous quality control (QC)[8,9].

qMR studies range from single-centre research on phantoms or healthy volunteers, through single-centre patient studies that aim to answer a specific clinical question, to multicentre randomised clinical trials of patients with more established imaging biomarkers. Earlier in this spectrum, the qMR IB is often the focus of the investigation, whereas later it is the tool used to investigate the disease state or investigational drug. It is in the propagation to a multicentre setting that more bespoke and novel qMR IBs [10] struggle to move from academic research tools to widely accepted techniques with clinical utility. To improve translation of qMR IBs, it is generally agreed that quality assurance (QA) procedures using specifically designed phantoms or test patients need to be performed to ensure the results of qMR are accurate and precise.

In this manuscript we provide an overview of current quality management procedures (including both QA and QC) for qMR IBs in a single country, namely the United Kingdom (UK). For clarity we define QA and QC below, derived from the ISO 9000 standard [11]:

QA – "Planned and systematic activities conducted to ensure that processes are performed – and data are generated, documented, analysed and reported – in compliance with the protocol, standard operating procedures (SOPs), good practice (GxP) and any other applicable regulatory requirements". QA covers activities such as assessment of scanners and ancillary equipment, regular phantom scans, imaging manuals and SOPs to define the scanning process.

QC - "Periodic operational checks to verify that data are generated, collected, handled, analysed, and reported according to protocol, SOPs, GxP and any other applicable regulatory requirements". QC includes phantom scans to identify scanner issues, checks of protocol compliance, the review of raw and analysed imaging data.

In this paper we define **Academic** qMR IBs as those that have demonstrated potential in research studies but are yet to be translated into clinical practice or used for decision-making in late-stage clinical trials. We focus on these in the context of qMR IBs that have been or are close to being successfully translated, either by being absorbed into clinical practice via community consensus (**Community** qMR IBs) or by taking a proprietary route via companies supporting qMR IBs with associated intellectual property (**Commercial** qMR IBs). See (Fig. 1).

As healthcare systems and the management of MR equipment differ significantly between countries, we aim to explore the current landscape for qMR IB use and quality management at research centres within a single territory. We focus here on the UK. In the UK most MR systems are either owned/operated by researchers, or owned by healthcare organisations; the former are used almost entirely for academic research, while the latter are primarily operated for routine diagnostic radiology by the National Health Service (NHS). This analysis is applicable to the UK only, but allows for a ready comparison with similar studies from other territories. For instance, studies by American Association of Physicists in Medicine (AAPM), Quantitative Imaging Biomarkers Alliance (QIBA) and Quantitative Imaging Network (QIN) [2,5,7,12–14] in the USA and Italian Association of Medical Physics in Italy (AIFM) [15–17].

We use a gap analysis to identify the key factors which require

improvement to allow academic qMR IBs to fulfil their potential and improve diagnosis and healthcare outcomes for patients. Here we use three complementary approaches to provide an overview of current qMR IB use and quality management practices for the first time.

2. Aim of the article

This overview aims to: (i) describe the UK landscape of qMR IB use in clinical research studies; (ii) provide a snapshot of QA/QC practices within the UK-based clinical research MR community; (iii) summarise current attempts to standardise and translate qMR IBs; and (iv) outline the ideal quality management practices and provide a gap analysis between current practice and this metrological standard. From this assessment of current practice, we aim to build upon principles outlined in Cancer Research UK (CRUK) and European Organisation for Research and Treatment of Cancer (EORTC) Imaging Biomarker Roadmap [18] and work towards consensus guidelines with relevant stakeholders to develop the level of consistency in data acquisition and analysis required to accelerate the translation of qMR IBs into clinical practice.

3. Literature and clinical research database search of quantitative MRI clinical research studies in the UK

The following methods were used to assess the current landscape of clinical research using qMR IBs and the associated quality management practices: 1) a literature search of published research in the past 10 years, split into two consecutive 5-year periods (01 Jan 2011–31 Dec 2015 and 01 Jan 2016–31 Dec 2020); 2) a database search of the use of qMR in clinical research studies in the second of those 5-year periods (01 Jan 2016–31 Dec 2020).

3.1. Method

3.1.1. Literature search

We conducted a literature search using PubMed for publications of clinical qMR research performed by UK-affiliated authors. The search covered all publications referencing MRI and MR spectroscopy (MRS) and at least one of the associated qMR methodology terms in the title or abstract, as listed in Table S1 (Supplementary Material).

Filters were applied to limit the search results to all Journal Articles describing studies conducted on humans (i.e. excluding preclinical studies), having at least one UK-based author and published within either of these two five-year periods: 2011–2015 and 2016–2020, inclusive. The search was also restricted to three publication types [19] – clinical trials, multicentre studies and validation studies, all of which are subsets of journal articles. Note that journal articles can be in multiple publication subtypes, or in none of the publication subtypes. A subsequent analysis was performed on the search results to determine the proportion of publications that included search terms associated with quality management within their title or abstract body. See Table S1.

3.1.2. Database search

We conducted a database search for all clinical research studies using qMR, filtered down to studies from 2016 to 2020 with at least one UK site, using the following three freely accessible and searchable online sources:

- the National Institute for Health Research Clinical Research Network (NIHR CRN) portfolio;
- the ClinicalTrials.gov database;
- the ISRCTN registry (originally this stood for International Standard Randomised Controlled Trial Number. The scope has now changed but the abbreviation remains).

The search covered studies referencing terms in the title or abstract as listed in Table S3 (Supplementary Material). Search results were sorted as shown in Figure S3. The data were again analysed to determine the proportion of publications that included search terms associated with quality management within their title or abstract body.

3.2. Results

3.2.1. UK literature search

In 2011–2015, 8.1 % of all journal articles mentioning qMR were clinical trials, 4.7 % were multicentre studies and 2.1 % were validation studies. All journal and all 3 sub-types of publications have increased in number between the periods 2011–2015 and 2016–2020, with the largest increases being for multicentre studies (112 %) and clinical trials (66 %), and only a modest increase in validation studies (11 %). In total, clinical trials are the most common subtype of publication (7.7 %, 2011–2020 inclusive), followed by multicentre studies (5.3 %) and validation studies (1.6 %). See Table 1 for details.

The number of publications where quality management terms are included also increased for all journal articles between the periods 2011–2015 and 2016–2020. As we would expect, the majority of validation studies (nearly 90.0 %) include quality management terms within their title or abstract. Between 2016 and 2020, only 27.7 % of all journal articles and 16.0 % of clinical trials included quality management terms in the title or abstract. However, 34.4 % of multicentre studies mentioned quality management terms, a 207 % increase in studies since 2011–2015.

The literature search shows that the number of publications increased between the periods 2011–2015 and 2016–2020 for almost all qMR methods (Fig. 2a). Only a sub-section of the hyperpolarised MR methods - helium-3 imaging - reduced in number, primarily due to a decrease in the availability of the required gas [20] and a subsequent shift towards xenon-129 imaging. The most popular qMR method in the literature was Diffusion Tensor Imaging (DTI) (17.4 %), followed by volumetry, a measurement of volume using imaging such as voxel based morphometry.

Of the literature published between 2016 and 2020, the journal

Table 1
Number and percentage of journal articles and subtypes for a PubMed search from 2011 to 2015 and 2016–2020 for UK-based qMR literature, and those which include quality management terms.

	2011–2015	2016–2020	% increase in number from 2011 to 2015
All qMR literature			
All journal articles (JA)	1265	2235	77 %
Clinical trials (CT)	102 (8.1 % of JA)	169 (7.6 % of JA)	66 %
Multicentre studies (MC)	59 (4.7 % of JA)	125 (5.6 % of JA)	112 %
Validation studies (VS)	27 (2.1 % of JA)	30 (1.3 % of JA)	11 %
Including quality management terms			
All journal articles (JA)	312 (24.7 % of JA)	620 (27.7 % of JA)	99 %
Clinical trials (CT)	18 (17.6 % of all CT)	27 (16.0 % of all CT)	50 %
Multicentre studies (MC)	14 (23.7 % of all MC)	43 (34.4 % of all MC)	207 %
Validation studies (VS)	24 (88.9 % of all VS)	27 (90.0 % of all VS)	13 %

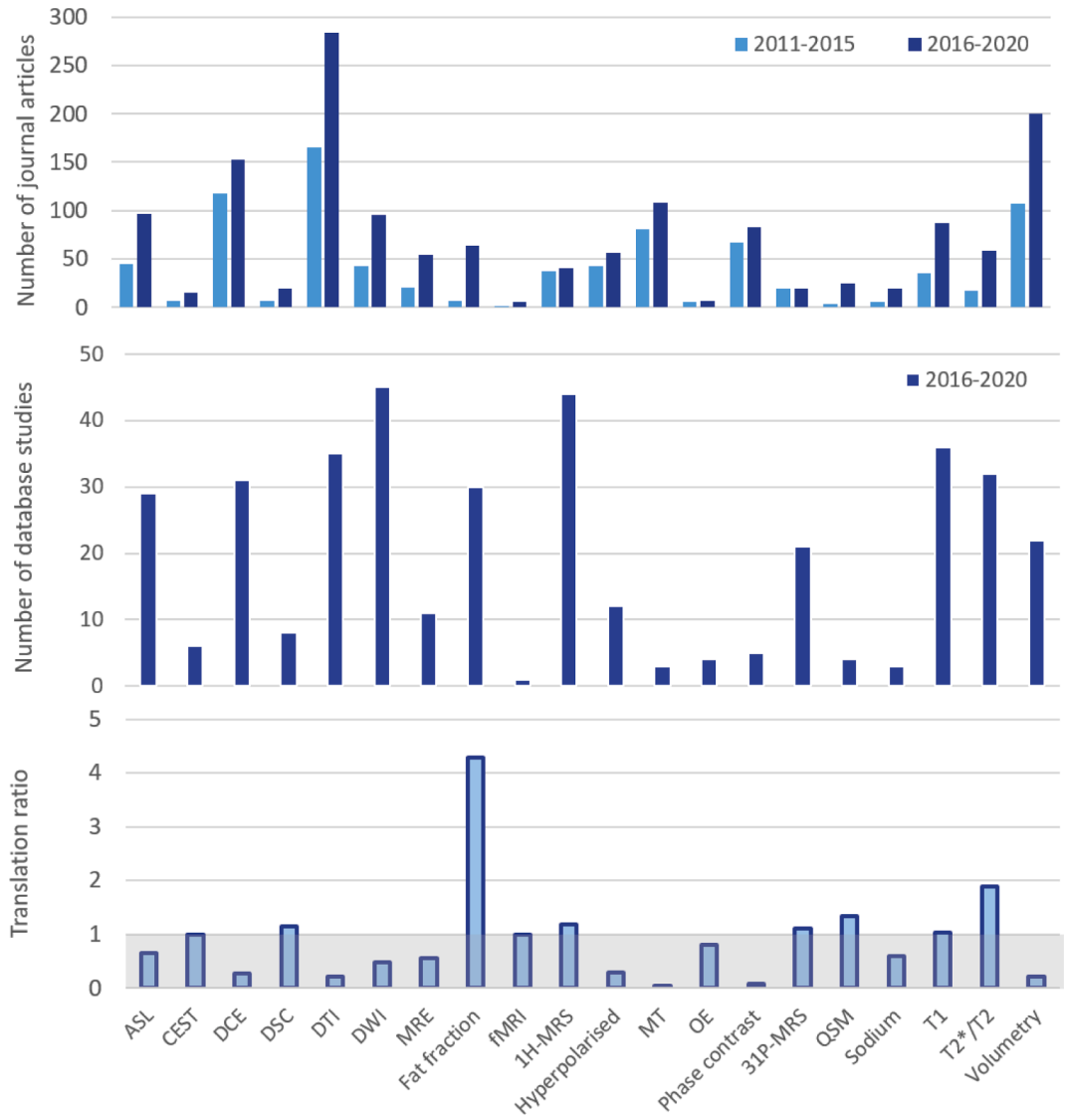


Fig. 2. Summary of results from the literature and clinical research database search a) Number of journal articles using quantitative MR methods during 2011–2015 and 2016–2020; b) Number of clinical research database studies using quantitative MR methods 2016–2020; c) Ratio of clinical research database studies in 2016–2020 to journal articles published in 2011–2015 for each quantitative MR method - the "translation ratio". Abbreviations: arterial spin labelling (ASL), chemical exchange saturation transfer (CEST), dynamic contrast-enhanced (DCE), dynamic susceptibility contrast (DSC), diffusion tensor imaging (DTI), diffusion weighted imaging (DWI), magnetic resonance elastography (MRE), functional MRI (fMRI), magnetisation transfer (MT), oxygen-enhanced (OE), quantitative susceptibility mapping (QSM).

articles that most frequently included quality management terms as a proportion of all journal articles involved dynamic contrast-enhanced MRI (DCE, 14.0 %) and DTI (8.8 %). However, when calculating the percentage of literature published including quality management terms normalised to the total number of journal articles per qMR method, we find that methods such as Chemical Exchange Saturation Transfer (CEST, 60 %), sodium imaging (58 %) and MR elastography (MRE, 41 %) have a greater proportion of studies referring to quality in the title or abstract. This may reflect that these methods are less well established and more technical validation of the IBs is being performed or that there are more developed quality management practices within those fields. See Table S2 and Figure S2 in the Supplementary Material for details.

3.2.2. Database search

We identified 248 unique clinical research studies using qMR imaging conducted in the UK with study start dates from 2016 to 2020, and extracted information about the MR method used, disease category, funder type, study type (observational or interventional), location and number of sites and number of subjects.

For the clinical research studies on the database that used qMR imaging, the median number of subjects was 56. Only 31 % of studies had >100 subjects and 56 % of those studies with > 100 subjects were multicentre. However, in total, 29 % of the recorded clinical research studies using qMR IBs were multicentre studies, with only 12 % having >8 sites. The database search revealed 72 multicentre studies, of which 40 were in neuroimaging (55.5 %), 6 in cardiovascular imaging (8.3 %), 5 in cancer imaging (6.9 %), and 21 in "other" (29.2 %). See Figure S4 for details.

Industry funded 21 % of the registered studies, 60 % of which were multicentre. 64 % of multicentre studies with over 8 sites had industry sponsors, underlining the increasing enthusiasm from industry as clinical qMR IBs move through the translational pathway. Of the multicentre studies, 43 % were based in the UK only and 34 % of multinational studies were led by UK institutions, highlighting the need for international conventions and standards of practice.

When examining the number of clinical research studies carried out

by each institution, we see that most studies, particularly multicentre studies, are led by large hospitals or universities. There is also a gradient in the number of studies performed, from universities to large hospitals and to smaller sites (Figure S5).

Fig. 2b shows that DWI is the most commonly used technique according to the clinical studies database search (45 studies, 18 %), followed by ^{1}H-MRS (44 studies, 18 %), T_1 (36 studies, 15 %) and DTI (35 studies, 14 %). Similar trends are apparent in both literature and clinical research database searches for the diffusion methods; however, Fig. 2c shows that the journal articles using ^{1}H-MRS and T_1 methods during 2011–2015 translate to an almost equal number of database studies in 2016–2020 (ratio $\approx$ 1). The "translation ratio" is even higher for quantitative susceptibility mapping (QSM), T_2/T_2^* and fat fraction (ratio $>$ 1).

The database revealed only 6.5 % of clinical research studies included quality management terms in the abstract. This is a lower percentage than that observed in the literature (>25 %), which may reflect the format requirements of the database entry or a lack of appreciation by the imaging community for the importance of describing what quality management procedures were used in the study.

More details of the database search can be found in the Supplementary Material.

3.3. Limitations

There were some limitations to the search method. Firstly, all possible variations of the name of the qMR IB may not have been included in the literature/database searches. Studies may have been included incorrectly due to, for example, the search terms including literature with the same name or acronym, the inclusion of a qMR term in the title or abstract that was not used in the study, the inability to distinguish multi-parametric studies from studies using separate qMR methods. Only quantitative MR method search terms were used, so in cases where the qualitative use of the biomarker is common, quantitative studies may have been excluded if terms such as "quantitative fMRI" or "quantitative T_1" or "apparent diffusion coefficient" were not included in the title or abstract. Relevant studies may also have been excluded from the results if they did not use the terms "MRI" or "MR" in the title or abstract.

Although PubMed is an extensive database, it does not contain the entirety of published work. Any journal articles that were not published in one of PubMed's indexed journals, or un-published studies that may have failed due to inadequate QA, have been excluded from this review.

The searches were intended to include only studies where there was a UK affiliation; however, the inclusion of a UK-based author would not necessarily mean the study is UK based. Additionally, details of the quality management process may not be found in the title or abstract, as in some instances they justify the robustness of the results rather than being the purpose of the research. Lastly, it is conceivable that active quality management processes may have been overlooked entirely in some instances by authors, either due to perceived low interest for publication or due to oversight. As such, some aspects of quality management will have been missed.

In the database search, the information we required for the study to be included may be missing from the limited database entry, particularly with regards to quality management. Additionally, not all clinical research studies will have been registered on a database.

4. A survey of national clinical research MRI and quality management

A survey of research institutions across the UK was designed to ascertain the current status of clinical MRI scanner equipment availability, QA/QC practices and quality management procedures.

4.1. Method

Representatives from all UK research institutions and centres conducting clinical MR IB research studies were invited, via a number of mailing lists (British and Irish Chapter of International Society for Magnetic Resonance in Medicine (ISMRM) and MRIPHYSICS), personal invitation and the National Cancer Imaging Translational Accelerator (NCITA) website (https://ncita.org.uk/ncita-launches-national-clinical-mri-quality-assurance-and-quality-control-survey), to complete a survey launched by NCITA to ascertain the current QA/QC practices and quality management procedures used for human MRI scanners and associated equipment at research institutions across the UK. The survey was created in SelectSurvey, an online tool for creating surveys and questionnaires. The survey used a series of multiple-choice questions and free-text boxes and was open from 25th August 2020 to 4th October 2020. See Supplementary Material for a copy of the survey and a summary of the results for each survey question (Table S8 and Figure S6-S24).

4.2. Results

4.2.1. Demographic

Data were collated from 33 respondents at 31 institutions based at universities (17, 52 %), NHS hospital trusts/boards (13, 39 %), the private sector (2, 6 %) and charities (1, 3 %). The respondents were principally imaging scientists in either a clinical (42 %) or a research (26 %) role, radiographers (26 %) or imaging facility staff members (6 %). Significant research was performed in neuroimaging at 85 % of the institutions, cancer at 73 %, cardiovascular at 48 % and musculoskeletal at 52 %.

4.2.2. Scanners and ancillary equipment

The MR scanners most commonly available for clinical research were manufactured by Siemens Healthineers (61 %), followed by GE Healthcare (21 %) and Philips (18 %). Scanners of magnetic field strengths of 3 T (53 %) and 1.5 T (44 %) were dominant, with 3 % of scanners being 7 T. Only 10 % of the MRI scanners for which responses were received were dedicated exclusively to research studies. Details were given for a total of 95 scanners, of which around 50 % were at least 5 years old and around 15 % over 10 years old. The aging imaging equipment in the NHS has recently been the focus of media reports and is emphasised in the NHS Long Term Plan [21]. In part, the establishment of National Imaging Networks, as part of the national imaging strategy, aims to utilise the collective buying power of the networks to update aging equipment [22,23]. This is exemplified by the average age of the scanners increasing with increasing clinical use: research only (5.0 years), mainly research (5.4 years) and mainly clinical (7.3 years).

4.2.3. QA/QC of the data acquisition

Nearly all the institutions completing the survey (91 %) had research agreements on all (73 %) or some (18 %) of their scanners, and the majority (85 %) had an on-site MR physicist to assist in the implementation and development of imaging protocols. Access to the clinical research MRI scanners to perform QA phantom scans was reported as sufficient by almost all respondents (97 %). This number perhaps does not reflect the reality in a smaller hospital setting, where there may only be very limited time available to scan phantoms. It may have been more pertinent to ask if sites felt able to increase the amount of QA performed on their scanners if the acquisition of qMR IBs required this.

Vendor-supplied phantoms were the most commonly used phantoms for QA purposes. These phantoms are supplied with the scanner and used for acceptance testing and regular QA procedures. Radiographers have access to vendor-supplied test programs that can be used to assess the signal characteristics and geometric accuracy on a regular basis. Most institutions had access to the American College of Radiology (ACR) phantom (76 %) to assess signal-to-noise (SNR) and geometric distortion

and many had access to the quantitative EUROSPIN phantom (45 %) – see Fig. 3a. This trend was observed over all phantoms available at the institutions, with the majority being SNR/geometric phantoms (63 %) and the minority quantitative phantoms (35 %) – see Fig. 3b. Most phantoms were used for QA on a monthly basis (30 %), although 44 % of institutions performed daily QA with at least one of their phantoms. It should be noted that a significant number of institutions tested only the head coil (43 %) or only the body coil (11 %) during QA, perhaps meaning the quality of QA for a study using a non-tested coil could be sub-optimal.

Data were transferred offsite using a range of methods, with 94 % of institutions using some kind of electronic/network/cloud-based transfer and 59 % at least in part relying on removable media, such as DVDs and pen/hard drives. Local records were generally printed out, filled in and stored either as hard copy (33 %) or scanned and stored electronically (37 %). The removal of personal information for research imaging data is an integral part of General Data Protection Regulation (GDPR) compliance and pseudo-anonymisation of the data is principally performed either on acquisition (43 %), likely through the use of a linked subject ID for research studies, on-site using anonymisation software (39 %) or on upload (18 %).

When asked about the QA procedures performed, most institutions reported acquiring vendor QA metrics (84 %), with 55 % acquiring quantitative MR metrics using third party or home-built phantoms. In terms of documentation, the majority of institutions had site-specific SOPs detailing equipment use, maintenance and QA practices (79 %); however, only 36 % used a quality management system (QMS) to define, improve and control processes [24] – see Fig. 3c. After scanner maintenance or an upgrade, no qualitative nor quantitative phantom/volunteer scanning was performed at 39 % of sites. Where post-maintenance/upgrade QA was performed, qualitative or quantitative scanning was performed equally at each institution.

Breaking down into the different quantitative MR methods, we find that DWI was performed at most of the institutions (85 % total, 15 % qualitative and 70 % quantitative) and oxygen-enhanced (OE) MR at the fewest (12 %, all quantitative) – see Fig. 4a. Quantitative relaxometry was performed at fewer sites than DWI, DTI or functional MRI (fMRI). However, it should be noted that fMRI analysis is generally not quantitative in nature and these figures likely reflect the use of qualitative threshold-based fMRI analysis methods. The popularity of these methods likely reflects activity in neuroimaging compared with cardiovascular and musculoskeletal imaging (Figure S8), where relaxometry is more common. This also agrees with the findings of the literature and database in Fig. 2.

In institutions where data for a given quantitative MR method were acquired, QA – defined as visual checks, phantom and healthy volunteer scans – was performed on average 69 % of the time (Fig. 4b). Spectroscopic, diffusion and relaxometry methods were more likely to undergo QA checks and this likely reflects the availability of suitable phantoms for these qMR methods. The time when institutions performed QA during a study varies. On average, 58 % of the institutions performed QA only at the beginning of the study, as opposed to throughout the study.

4.2.4. QA/QC of the acquired data and analysis

Quality management of the entire pipeline – from protocol implementation and testing, through data acquisition to data analysis – is key to maintaining good quality in a study. Quality control of the acquired data involves ensuring that data are received, that there were no protocol deviations, image quality is adequate, regions of interest (ROIs) are appropriately defined and data analysis is valid and implemented according to the SOPs. Key to this is the development and management of suitable data analysis methods and code review procedures. Journals increasingly require data and/or code sharing as a prerequisite for publication [25] and reproducible research has recently been a key topic for the MR community [26].

The survey found that QC checks of the acquired data were generally underdeveloped, with only 67 % of institutions having SOPs for the analysis of results, 40 % performing data quality checks and 33 % protocol deviation checks on the acquired data. Institutions use a broad range of methods to analyse data, with vendor's software and in-house code being used at 97 % and 90 % of the institutions, respectively. However, although in-house developed code was used at the majority of institutions, remarkably only 21 % of institutions had any code review or software QMS in place for in-house data analysis code. See Fig. 5.

Overall, the survey revealed that the earlier sections of the QA/QC pipeline - i.e. QA using vendor geometric/SNR phantoms and SOPs for equipment use, maintenance and QA - are generally well performed by institutions carrying out clinical MR research. However, QA for qMR IBs is less commonly performed and the QC of the acquired data and data analysis was underdeveloped, with a lack of consistency across institutions.

4.3. Limitations

There were some limitations to the data from the survey. Firstly, the responders were asked to complete the survey via a number of mailing lists and by personal invitation, therefore the sample may not be fully representative of the wider UK imaging community. Secondly, not all questions were answered in full. Typically, the questions at the beginning of the survey were completed by all responders, whereas the latter questions were not answered fully. However, this disparity was taken into account in the analysis and does not impact our conclusions.

5. Discussion

5.1. Linking the literature and database searches with the survey results

We have used three complementary methods to assess the status of QA and QC in qMR in the UK. The methods are: 1) the number of qMR journal articles published; 2) the use of the qMR method in clinical research studies registered to a database; and 3) the reported use of qMR via a national survey.

The more popular qMR methods from the survey tended to have a high number of associated journal articles and clinical research studies. Not all journal articles and clinical research studies used a single qMR method. From the literature search we see that 1.6 % of all journal articles from 2011 to 2015 used the term "multi-parametric" MR and this increased to 3.9 % by 2016–2020. The database results from 2016 to 2020 found 38 % of registered studies used more than one qMR method, suggesting an increase in the combination of qMR methods for clinical research. The translational potential of this has been shown with the recent NICE recommendation for a multi-parametric MRI scan as a first-line investigation for people with suspected clinically localised prostate cancer [27].

Instead of looking at how many journal articles we can get out of a research study, here we look at how many clinical research studies the corresponding literature produces. The ratio of the number of clinical database studies carried out (2016–2020) to journal articles published (2011–2015) tells us something about translation during that time period. Fig. 2c shows that qMR methods such as fat fraction, QSM, DSC, MRS (both ^{1}H and ^{31}P), $T_1/T_2/T_2^*$ calculation and CEST have a higher "translation ratio" per journal article than more commonly used methods such as DWI, DTI and phase contrast.

There is also a discrepancy between the percentage of journal articles that include quality management terms and the number of institutions that report using QA via the survey. Quantitative T_1, DWI and MRS all have high levels of reported QA from the survey, perhaps due to easier access to phantoms (i.e. Eurospin, NIST diffusion, in-house phantoms). By contrast, the percentage of published journal articles per qMR method that refer to quality management terms in their title or abstract is relatively low, suggesting that many users of qMR methods that acquire QA data do not necessarily report this information in the abstract,

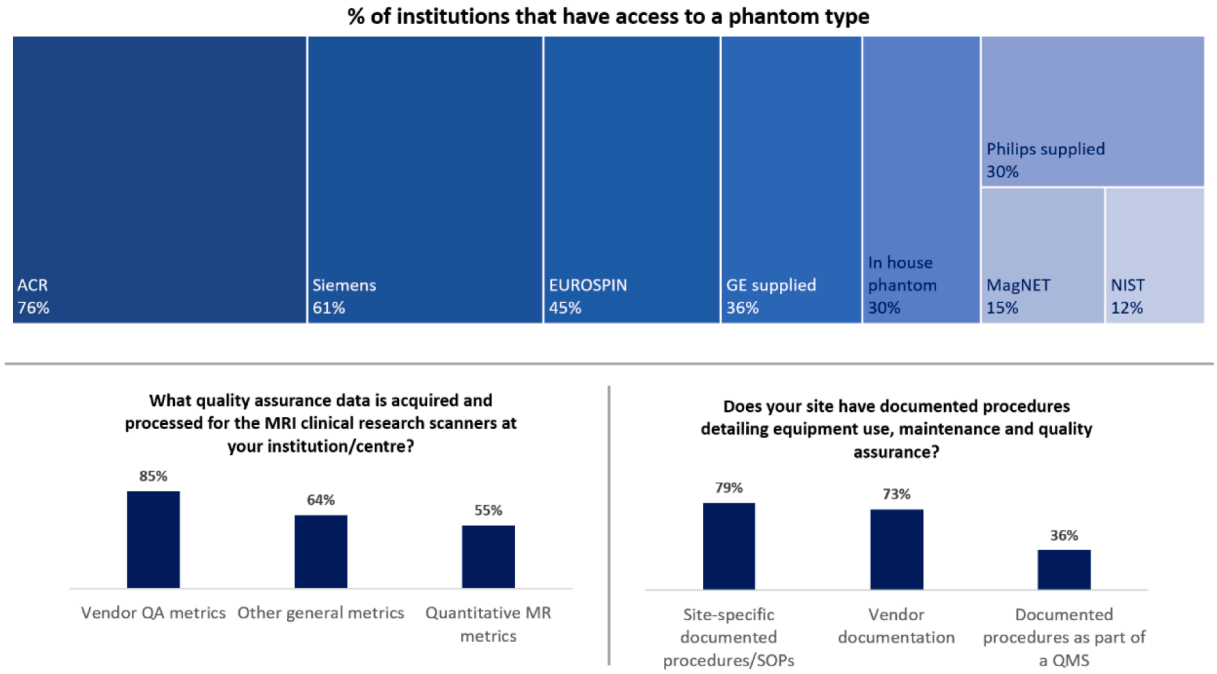


Fig. 3. Phantom QA survey questions: a) the percentage of institutions that have access to a given phantom; b) the percentage of institutions acquiring and processing different types of QA data; c) the percentage of institutions using different types of documented procedures for equipment use, maintenance and QA.

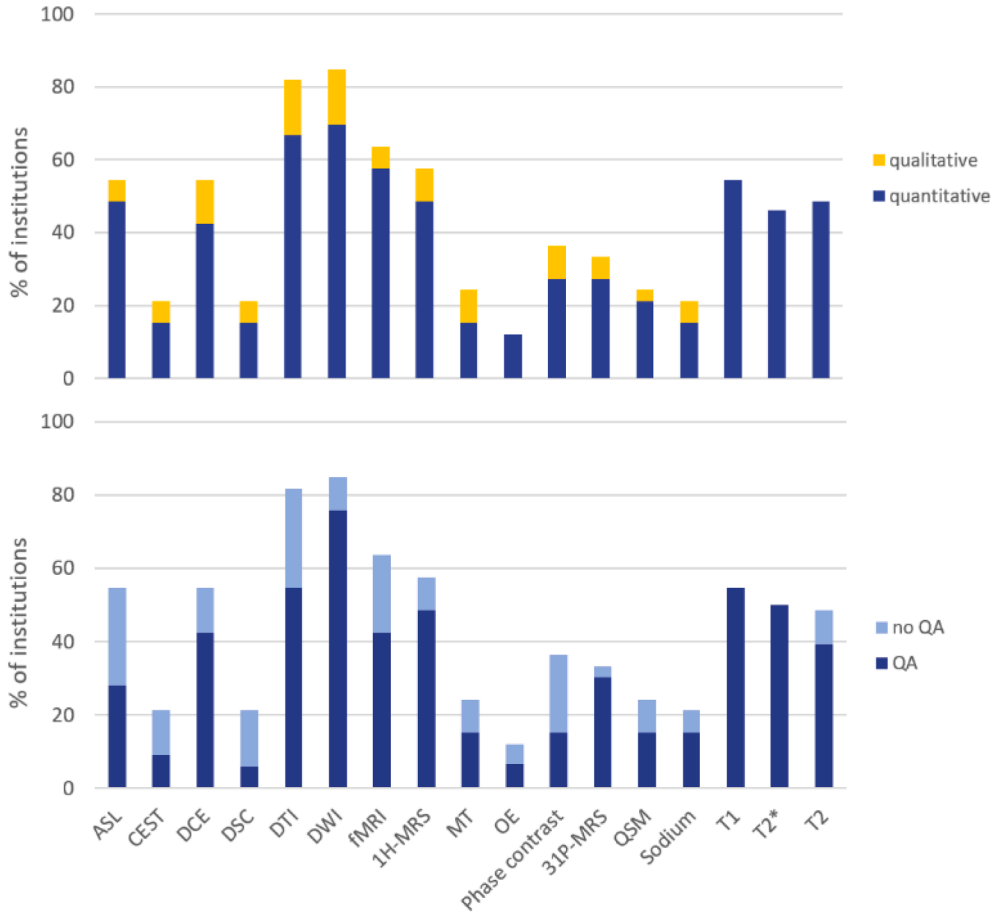


Fig. 4. Quantitative MR method survey questions: a) the percentage of institutions using an MR method, quantitatively and only qualitatively; and b) the percentage of institutions performing QA per MR method.

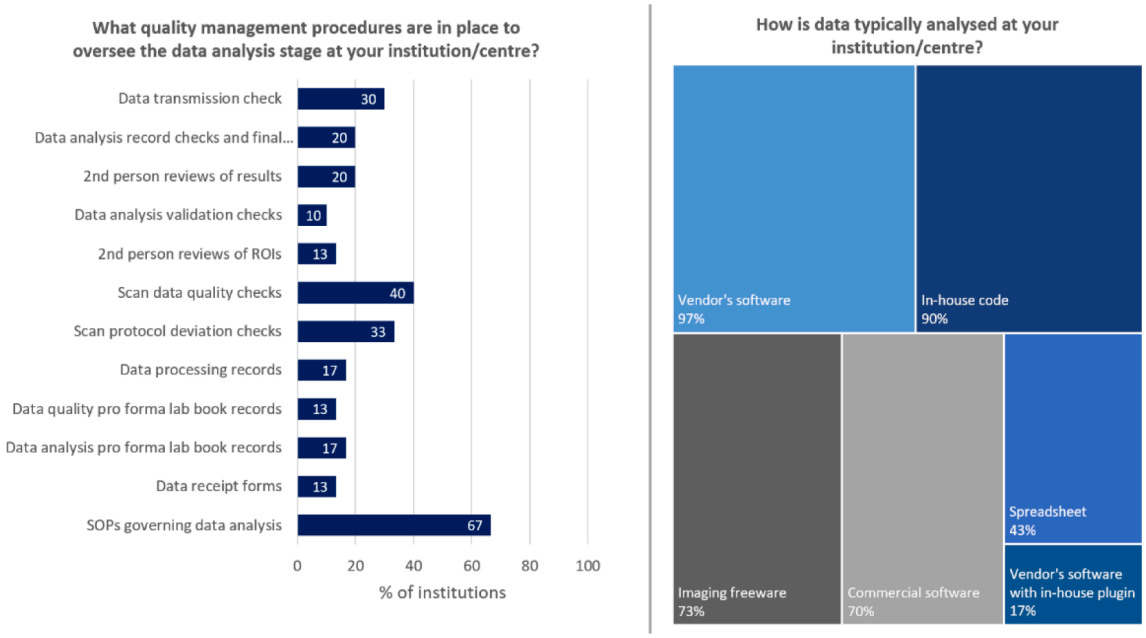


Fig. 5. Data and analysis QC survey questions: a) quality management procedures for acquired data and data analysis; b) methods of data analysis used by institutions.

although it may be in the journal article elsewhere. However, for qMR methods such as CEST, DSC, phase contrast, QSM and sodium imaging, the survey reported QA matches more closely that observed in the literature.

5.2. Standardisation, harmonisation and quality management of qMR data acquisition and analysis

As identified by a recent report from the AAPM [7], standardisation is the key to the success of clinical trials with qMR IBs as primary endpoints and to amassing the evidence necessary to translate qMR IBs into multicentre use and clinical practice. The entire pipeline from imaging platform to data acquisition protocols, image post-processing and analysis, data transfer and reporting system needs to be considered. Clear quality management guidelines need to be established to allow qMR IBs to be assessed at multiple sites. Before activation of the trial or study, there needs to be established evidence of scanner qualification and/or accreditation, highly harmonised imaging protocols, test scans

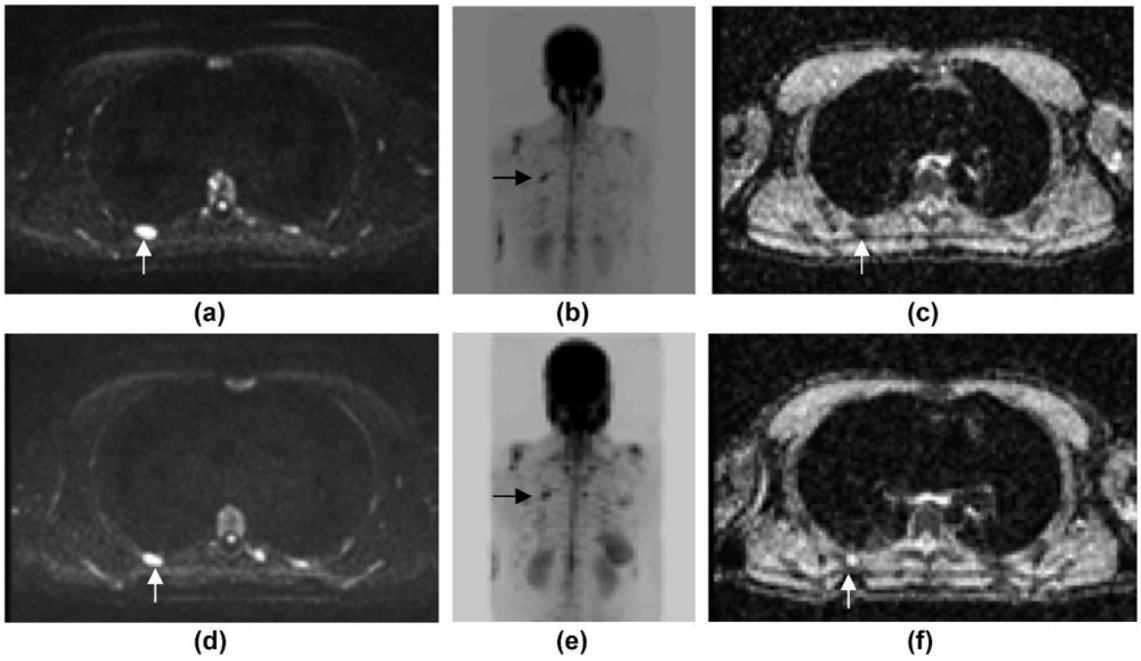


Fig. 6. Example of clinical translation of ADC from literature. Axial b = 900 s/mm^2 DWI, b = 900 s/mm^2 maximum intensity projection (MIP), and ADC map pre- (a-c) and post- (d-f) chemotherapy in a patient with myeloma. The right posterior rib lesion (indicated by the arrows) did not reduce in size and appears unchanged post-treatment on the b = 900 s/mm^2 and MIP images. Reprinted from Winfield *et al.* Clinical Radiology. 2021;76(10):715–27. [72], with permission from Elsevier. Copyright © 2021 The Authors.

Table 2
Gap analysis of quality management comparing current status to ideal activities defined, where appropriate, as the metrological standard.

Quality management domain			Why this is important?	Ideal quality management activities	Examples of QAQC practice elicited from survey/literature	Gaps and challenges
GENERIC METROLOGY CONSIDERATIONS						
Accuracy (lack of bias)			If qMR is inaccurate or biased: • Clinical decisions triggered when qMR IB > threshold cause patient harm • Drug approvals based on qMR IB > threshold in clinical trials place unsafe or ineffective drugs on the market • qMR findings in clinical trials wrongly stop or progress investigational drug developments	• MR scanning of well-characterised traceable phantoms with documented calibration chain to provide SI traceability for each measurand • Analysis of software phantoms • Biological validation against preclinical model or cell culture/ organoids.	• Most institutions reported that they had time to perform regular routine QC (97%). • 55% of institutions acquire quantitative metrics during routine QC. **SEE FIGURE S13 and FIGURE 3/S14**	• Well-characterised traceable phantoms do not exist for all qMR biomarkers • Very few phantoms provide SI traceability. Without this, variability cannot be measured • If phantoms are not traceable, no chain of calibration is available for reference values • Many standard tests of image quality are subjective and non-quantitative • Uncertain translation of phantom to human.
Precision (treated as stochastic)	**Repeatability**	**Same site, same scanner, same operator, same subject**	• Repeatability places a limit on the uncertainty of the measurement and therefore the confidence with which clinical or research decisions based on a single qMR measurement • Incorrect estimates of repeatability cause incorrect sample size estimates leading to futile exposure of patients in clinical trials.	• Repeated analysis of same scan • MR scanning of same subject on successive days (double-baseline studies) • Benchmarking of scanner and reconstruction by repeated scanning of well-characterised traceable phantom on successive days.	• Only 49% of institutions perform daily QA with at least one of their phantoms. **SEE BOX 1**	• Repeatability measurements may pose an unacceptable burden on human subjects e.g. if a contrast agent carries a risk of harm.
		Within site variation	• qMR use for monitoring, response and pharmacodynamic measurement • Treatment planning for radiotherapy often requires sequential scanning • IB is measured sequentially for a research or clinical decision so within site repeatability is critical • Different members of staff scan patients or prepare phantoms on different days.	• Regular, repeated MR scanning of well-characterised traceable phantoms within a site • Defined SOPs/QMS for QC and patient scanning.	• 79% of institutions have documented site-specific procedures and SOPs which help limit within site variation. • Only 36% of institutions have document procedures as part of Quality Management System (QMS). **SEE FIGURE 3**	• Institution-based or UK-wide QMS to manage quality of clinical research studies.
	Reproducibility	**Different scanners across multiple sites**	• qMR use as a screening, diagnostic, prognostic and predictive IB only measured once. • Irreproducibility places a limit on the consistency of the measurement across different sites and therefore the confidence in clinical or research decisions based on a single qMR measurement • Incorrect estimates of reproducibility cause incorrect sample size estimates leading to futile exposure of patients in multicentre clinical trials.	• Reproducible QA at multiple centres • MR scanning of well-characterised traceable phantoms • Travelling healthy or patient volunteers • Meta-analyses.	**SEE BOX 1**	• Reproducibility measurements may pose an unacceptable burden on human subjects e.g. if frail patients are requested to travel to different cites • Without traceability in test objects, variability between test objects (and thus reference values used as ground truths) cannot be quantified. • Institution-based or UK-wide QMS to manage quality of clinical research studies.

(*continued on next page*)

Table 2 (*continued*)

Quality management domain		Why this is important?	Ideal quality management activities	Examples of QAQC practice elicited from survey/literature	Gaps and challenges
Linearity		• May affect whether a patient is qMR IB positive vs qMR IB negative.	• MR scanning of well-characterised traceable phantoms.	• 100% of institutions that conducted quantitative T1 or T2 measurements reported carrying out QA tests using phantoms • 29% of institutions that conducted quantitative DSC measurements reported carrying out QA tests using phantoms **SEE FIGURE 4**	• Well-characterised traceable phantoms do not exist for all qMR biomarkers • Uncertain translation of phantom to human • Use of analytical procedures methodology.
		QAQC OF DATA ACQUISITION			
Geometric	**Dimensional**	• In the case of IBs which are extensive variables (*e.g. volume*), while dimensional accuracy of routine-use clinical MR is quite good it may not be adequate where very small/slow changes are clinically significant • In the case of qMR IBs which are intensive variables (*e.g.* K^{trans}), dimensional inaccuracy of ROIs may propagate to qMR IB inaccuracy *[e.g. brain, cartilage]* • Particularly important when images require registration in multi-parametric or multi-modality studies.	• MR scanning of well-characterised traceable phantoms • Quantification of spatial distortion and uncertainties due to distortion correction • Independent verification of voxel size.	• 100% of institutions have access to an SNR/geometric phantom • 71% of institutions have access to the ACR MRI accreditation phantom. • Only 44% of institutions perform routine QA using both a head and body coil, most just do one or the other. **SEE FIGURE 3/S15**	• Well-characterised traceable phantoms do not exist for all qMR biomarkers • Uncertain translation of phantom to human.
	Gradient and B_1 accuracy	• Some qMR IBs are extremely vulnerable to inaccurate gradient amplitudes, especially in settings where gradient accuracy varies spatially *[e.g. ADC]* • Some qMR IBs are very vulnerable to inaccurate RF amplitudes, especially in settings where B_1 homogeneity varies spatially *[e.g. T_1 measured by VFA]*.	• MR scanning of well-characterised traceable phantoms • Quantification of gradient non-linearity.		
SNR		• SNR places a limit on qMR IB repeatability and limits of quantification, especially for small structures *[e.g. mets, atheromas]*.	• MR scanning of well-characterised traceable phantoms • Analyses of propagation of errors through entire measurement pipeline, including analysis.		• Well-characterised traceable phantoms do not exist for all qMR biomarkers • Error propagation dependent on coil design and image encoding • Some analysis methods (e.g. machine learning) do not currently allow uncertainty propagation.
Motion		• Motion-induced qMR IB inaccuracy can be mitigated by motion compensation strategies and algorithms, but different mitigations may have different effects, leading to irreproducibility *[e.g. lung T_1]*.	• MR scanning of well-characterised traceable phantoms • MR scanning of software phantoms • Analyses of propagation of errors • Analysis of sensitivity to scan and object parameter uncertainty using software phantoms.	• Only 2% of institutions have access to a dynamic/motion phantom. **SEE FIGURE S15**	• Well-characterised traceable phantoms do not exist for all qMR biomarkers • Consensus guidelines lacking • Uncertain translation of phantom to human • Some analysis methods (e.g. machine learning) do not currently allow uncertainties to be estimated.
IB-specific quantification		• Validation against underlining biology is necessary to ensure qMR IBs are describing proposed physiology • Need to know that the range of a qMR IB measured on a given scanner is sufficient to quantify change observed in subjects.	• NIST-ISMRM working together to produce a series of traceable well-defined phantoms for qMR • Use of Bradford-Hill criteria to prove biological validity of IBs, as	• 45% of institutions have access to the EUROSPIN phantom • 12% of institutions have access to the NIST system or system lite phantom.	• Well-characterised traceable phantoms do not exist for all qMR biomarkers • Consensus guidelines lacking • Uncertain translation of phantom to human.

(*continued on next page*)

Table 2 (*continued*)

Quality management domain	Why this is important?	Ideal quality management activities	Examples of QAQC practice elicited from survey/literature	Gaps and challenges
		suggested in Cancer Imaging Biomarker roadmap [18].	• 30% of institutions rely on in-house phantoms, which may be less traceable and reproducible than commercially supplied phantoms. **SEE FIGURE 3/S15**	
		QAQC OF ACQUIRED DATA		
Imaging protocol compliance	• qMR IBs typically require data to be acquired with a specific imaging protocol to enable quantitative metrics to be calculated.	• Manual checks of specific DICOM headers across images • Automatic checks against DICOM header information.	• Only 33% of institutions perform protocol deviation checks. **SEE FIGURE 5/S21**	• Freely available protocol checker • Institution-based or UK-wide facility to manage quality of clinical research studies (i.e. a Core Lab).
Data quality	• Image artefacts and poor SNR within ROIs lead to incorrect calculation of the qMR IBs • Variation in image quality within sites or between sites may bias results or decrease sensitivity to change.	• Manual checks by radiographers and data analysts • Automatic methods (e.g.. deep learning).	• 73% of respondents would like access to centralised data review and quality control. • Only 40% of institutions perform data quality checks. **SEE FIGURE S25/S21**	• Automatic artefact recognition.
		QAQC of data analysis		
Region-of-interest definition	• Incorrectly placed ROIs impact the calculation of summary statistics from qMR that are frequently used to validate, qualify and use IBs • Dimensional inaccuracy of ROIs may propagate to qMR IB inaccuracy *[e.g. brain, cartilage]*.	• Suitably trained person defining ROIs using commercial software or imaging freeware. • Second checker desirable to validate ROI definition performed correctly	• Only 13% of institutions have a second person review of ROIs. **SEE FIGURE 5/S21**	• ROI tools integrated into data repositories to remove need for data transfer/external software • AI-guided ROI definition to aid reproducibility and reduce time.
Fitting procedure and algorithm	• Software used to calculate qMR parameters ranges from vendor or commercial software, community freeware to in-house code • The optimisation, implementation and choice of models affects the outcome and the ability to compare clinical research results from different research groups.	• Community-led code sharing initiatives: e.g. OSIPI, UKRIN, qMRLab • Use of institution-based or multi-centre code repositories *(e.g. GitLab, GitHub*) • Code review procedures.	• 67% of institutions have SOPs that govern data analysis. • Only 33% of institutions have any data analysis support. **SEE FIGURE 5/S21 and FIGURE S12**	• Support for research software engineers • Increased funding for community code sharing initiatives.
		SUSTAINABILITY AND TRANSPARENCY		
Evolution in MR hardware and software	• Vendors offer different data acquisition and analysis options for different qMR methods • Upgrades to MR hardware and software leave clinical research studies vulnerable to variation, particularly for multi-centre and/or long-term studies.	• Pre- and post- upgrade/ maintenance scanning • MR scanning of well-characterised traceable phantoms.	• 61% of institutions acquire quantitative QC data on a phantom pre- and post- upgrade/ maintenance. • Only 15% of institutions acquire quantitative QC data on a volunteer pre- and post- upgrade/ maintenance • The average age of MR scanners increased with increasing clinical use: research only (5 years), mainly clinical (7.3 years). **SEE FIGURE S17 and FIGURE S9**	• Well-characterised traceable phantoms do not exist for all extensive variables • Uncertain translation of phantom to human • Capabilities and age-discrepancy of clinical scanners must be acknowledged/addressed to aid translation of academic qMR IBs • Vendor engagement.

(*continued on next page*)

Table 2 (*continued*)

Quality management domain	Why this is important?	Ideal quality management activities	Examples of QAQC practice elicited from survey/literature	Gaps and challenges
Data formats	• Vendor specific data formats and conversion to non-DICOM data formats complicates data sharing and analysis as conversion tools that may are required and they may introduce errors • Private DICOM fields can differ between vendors and may be removed during anonymisation affecting the use of metadata.	• ISMRMRD (raw data) format (HDF5 files with XML header) • Adding qMR metadata to DICOM header (i.e. OSIPI lexicon group) • BIDS data structure (NIfTI with Json file with metadata).	• *Not covered in survey*	• Single data format unrealistic as metadata needs to evolve, but increased use and interoperability of data format standards and structures is possible.
Data sharing	• Sharing data between multiple research groups during and after studies aids collaboration and allows data to be checked, re-used and extended • Meta-analysis of large datasets acquired over time via multiple centres or studies • Supports reproducible research and improves trust in scientific endeavour.	• Data repositories such as The Cancer Imaging Archive, OpenNeuro and XNAT etc. • Use FAIR principles (Findable, Accessible, Interoperable and Reproducible).	• *Information about data sharing practices with the wider community were not obtained from the survey.* • 94% of institutions use an electronic/network/cloud services to transfer data. • 48.5% of institutions use the XNAT data repository to transfer and share data. • 18% of institutions use a non-encrypted hard-drive to transfer data, suggesting they may not have the infrastructure and support to share data effectively. **SEE FIGURE S21**	• Increased knowledge of data security, GDPR and privacy concerns • Better understanding of Intellectual property infringements • Community support to reduce fear of peers discovering errors or bad practice • Framework to support best practice and make sharing as easy as possible. • Increase accessibility and interoperability of data sharing tools.
Code sharing and software sustainability	• Data infrastructure needs to be persistent and should not disappear when individual project funding comes to an end. • Software also needs to be reviewed and updated to remain with current legislation and best practice.	• Emergence of generic, secure, cloud-based image databasing solutions (e.g. XNAT) • Community-led code sharing initiatives: e.g. OSIPI, UKRIN, qMRLab • Commercialisation • ISMRM MR Hub for code sharing • Use FAIR principles (Findable, Accessible, Interoperable and Reproducible).	• Only 21% of institutions have a software QMS or code review procedure in place for in house code. **SEE FIGURE S24**	• Best practice consensus required • No guidance on how data infrastructure best practice should be applied to individual projects • Standard process for containerising code (e.g. Docker or Singularity) for reproducible data analysis (can also be linked to repositories) • Funding for Research Software Engineers • Centralisation of national and international research infrastructure.
Open publishing and guidelines	• Sharing all data, analysis and methods of the study • Encourages reproducible research and best practice • Can report negative results leading to decreased publication bias • Increased collaboration and earlier feedback • Innovative peer-review • Open access publication increases worldwide access to scientific literature and increases visibility.	• Reporting guideline via Equator Network (www.equartor-network.org) - CONSORT, STARD, SPIRIT etc • MRM and MAGMA open science policy • MRM Highlights • Pre-registration and registered reports (peer-review before data collection and analysis) • Protocol preprints using protocol.io • eLab books.	• *Not covered in survey*	• Increased recognition of preprints • Full methodological information in publication, including all quality management procedures • Imaging-specific guidelines, such as those developed for AI [108].

on phantom and/or human subjects and a highly organised QA/QC program, which is then followed by competent and trained staff. Imaging equipment should be maintained and quality-assured throughout the whole study, re-qualified after hardware/software upgrades and following the detection of a problem data acquisition paused until the issue is resolved. Data should be excluded after QC failures and only included if correction is possible. Software quality management and code review are integral to the data analysis process. Independent central review of data quality and analysis may be used to avoid bias and to reduce measurement variability. Good quality management is key to supporting qMR IBs to move towards robust and implementable methods that benefit society.

Robust data acquisition and analysis are vital components of the pipeline, but almost all qMR methods both benefit and suffer from a plethora of possible variations in how data can be acquired and analysed. There have been concerted efforts by the qMR imaging community to come together and reach agreement on best practice with a range of consensus documents being published, primarily in the past five years. These range from qMR method-focused consensus such as recommendation on the use of arterial spin labelling (ASL) for clinical perfusion imaging [28] and DWI for whole-body imaging [29], to more specific applications of qMR methods in diseases such as cancer [30–33], cardiovascular disease [34], multiple sclerosis [35] and epilepsy [36], to organ specific applications such as the recent contributions of UK Renal Imaging Network (UKRIN) in kidney qMR [37–40].

Vendor-specific variation in data acquisition, storage and processing is problematic for standardisation of qMR IBs and their translation into clinical practice, but the use of consensus protocols, data sharing and good quality management practices can go a long way to reducing variability. Standardised traceable system phantoms are crucial for evaluating differences in qMR IB values across systems and over time [41] and have been utilised in multi-centre trials [42], and in the recent development of vendor-neutral pulse sequences (VENUS) [43].

Diligence in data acquisition must be mirrored by suitable care when designing, developing and sharing analysis code. Creating software that provides reproducible results and is sustainable is fundamental to developing and translating qMR IBs. Support for research software engineers by bodies such as the Software Sustainability Institute [44] and code repositories such as GitHub [45] and GitLab [46] are integral to maintaining good quality analysis tools long term, as is the ability to lock down code (with all dependencies) for the duration of a study via containerisation, using platforms such as Docker Hub [47] and Singularity [48]. Important initiatives such as the Open Science Initiative for Perfusion Imaging (OSIPI) [49], who are creating a software inventory for ASL, DCE and DSC code, and ISMRM's Reproducible Research Study Group's MR-Hub [50], which is a platform for researchers to share software with the rest of the community, are positive steps forward.

5.3. qMR imaging biomarker translation

5.3.1. Community initiatives

Many clinical bodies and groups of scientists/clinicians are increasingly focused on the standardisation and translation of qMR IBs for clinical application. Imaging biomarker translation requires a series of technical and biological validation projects which may be costly with low-priority for academic funders. Where a biomarker (imaging or otherwise) can be developed as a proprietary commercial product, then funds for validation can be raised from investors who will see a return when the diagnostic is approved and sold. Indeed, this approach has been successfully used for a number of qMR IBs, particularly in the liver [51–55]. However, many qMR IBs are difficult to commercialise in this way, so validation activities are frequently facilitated and/or funded by diverse stakeholders including professional associations (such as ACR and ISMRM), regulators (notably the Food and Drug Administration (FDA)), public–private partnerships (such as Innovative Medicine/Health Initiative (IMI/IHI)) and end-users (such as Pharma and CRUK).

The FDA offered guidance for clinical trial sponsors in a document on imaging endpoint process standards [4] which aimed to assist in the optimisation of imaging data quality from clinical trials supporting drug approval. The document encouraged the use of an "imaging charter" to give more detailed information on image acquisition, display, interpretation, and archiving than is generally found in the trial protocol. Moreover, FDA has established a biomarker qualification route [56] for validation of biomarkers which will be useful in multiple drug developments, but may not be commercially viable as diagnostics. Although designed for all biomarker modalities, imaging biomarkers have been prominent in this programme.

The National Cancer Institute (NCI) sponsored an open meeting during the 2008 ISMRM conference where one of the first qMR consensus papers was developed [33] and gave further support to the QIN who continue to work in the US to improve the role of quantitative imaging for clinical decision making in oncology [12,13,57].

Since 2007, the RSNA-sponsored QIBA has been working to bring researchers, clinicians and industry together to advance quantitative IBs in clinical trials and clinical practice. QIBA provides a range of profiles (organised collective results) and protocols (standardised imaging procedures) intended to increase the acceptance of quantitative imaging by the imaging community, clinical trials industry and regulators as proof of biological and pathophysiological change. There are currently qMR consensus profiles for DCE, DWI, and DSC for oncology, MRE of liver and MR-based cartilage compositional biomarkers [14].

ACR, together with ECOG-ACRIN cancer research group, maintain an Imaging Core Laboratory to support the group's imaging research activities [58] and a number of groups and subcommittees of the European Society of Radiology (ESR) merged to form the European Imaging Biomarker Alliance (EIBALL) [59,60], which aims to coordinate imaging biomarker activities in Europe and works alongside QIBA and others to provide an inventory of biomarkers validated for use in clinical trials [61].

The ISMRM study groups - Quantitative MR and Reproducible Research - have held workshops and published recommendations [62] with the aim of improving qMR IB translation. Other efforts include the development of a standardised QA protocol for DWI in multi-centre studies by the AIFM [15,16] and similar work for fMRI [63] and spectroscopy [64].

Examples of public–private partnerships include the Osteoarthritis Initiative (OAI) [65], Alzheimer's Disease Neuroimaging Initiative (ADNI) [66] and several IMI/IHI projects [67]. More recently, CRUK's National Cancer Imaging Translational Accelerator (NCITA) [68] is an initiative developed to promote and support the validation and translation of cancer IBs and develop UK-based infrastructure to support multicentre quantitative imaging trials, including the development of an MR Core Lab to support quality management of qMR IBs in multi-centre studies.

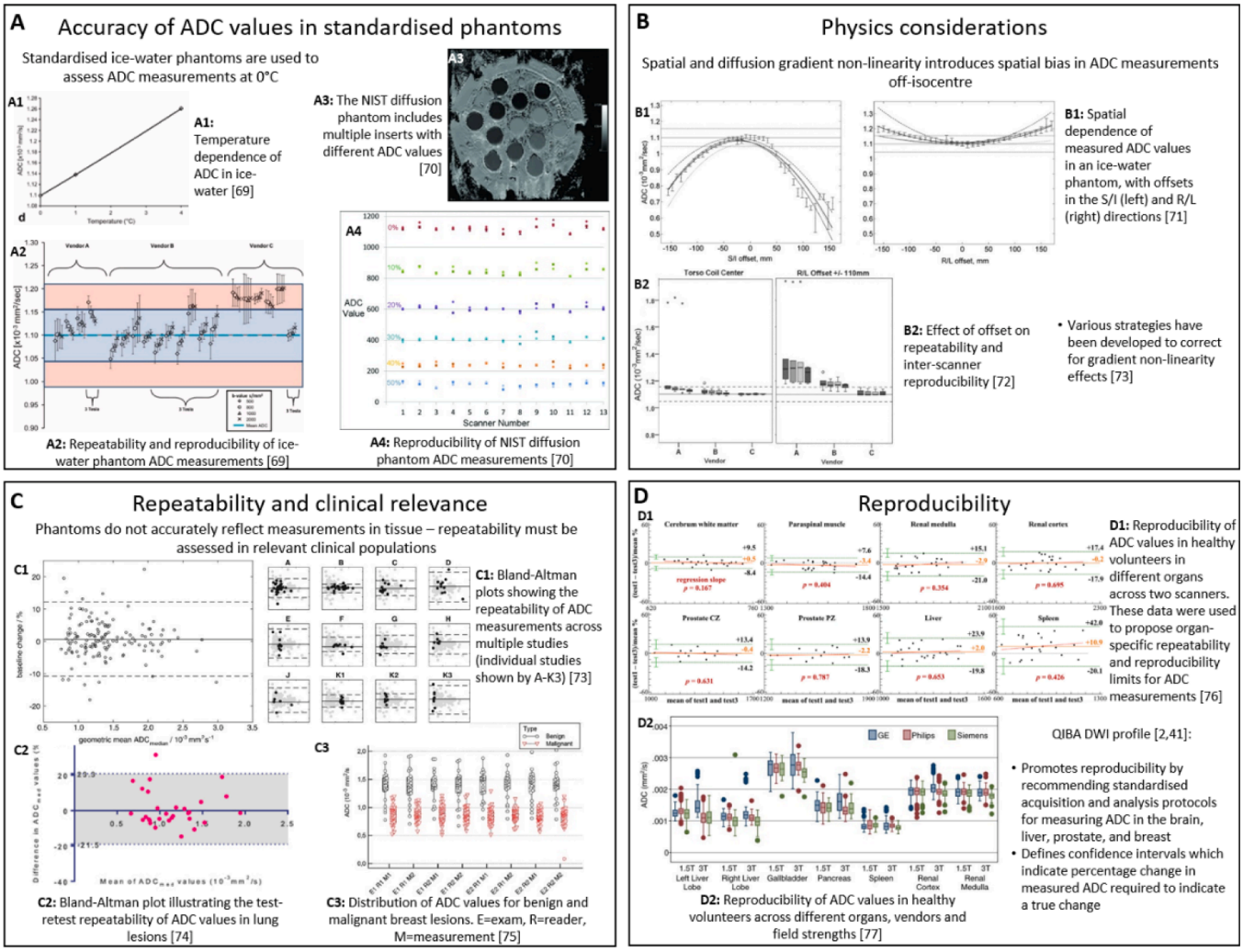
5.3.2. qMR IBs on the translational path

An academic qMR IB with a long track record that has gone some way towards translating into clinical use in oncology is the apparent diffusion coefficient (ADC) [33,69–71], measured using DWI. Fig. 6 shows an example of the clinical use, with an increase in lesion ADC estimates between pre- and post-treatment measurements in a patient with myeloma [72]. The high ADC in (f) confirms that the lesion has been treated and the patient does not have residual disease. This information guided decision-making as a patient with residual disease would require either further treatment or close imaging/biochemical surveillance. Box 1 describes the array of validation work that has been amassed by the DWI community to support this translation into clinical practice. However, decades of academic DWI studies show there is also potential for application-specific ADC thresholds to be defined [73–75] and for microstructural and membrane permeability measurements to be extracted via model fitting [76–79].

Although the acquisition of data to calculate ADC often suffers from a

lack of standardisation, it is a relatively simple qMR IB to calculate. ADC maps are also available via standard vendor software, without the need for offline processing. A more complex qMR IB such as the DCE-derived parameter, K^{trans}, is calculated by applying a pharmacokinetic model to a dynamic series of images acquired before, during, and after the injection of a paramagnetic contrast agent and has been used extensively in clinical trials of anti-vascular agents [80] and other settings, including cerebrovascular disease [81] and joint disease [82]. If we compare the evidence supporting this qMR IB with the evidence for ADC presented in Box 1, we find that there are many extra challenges for K^{trans}. Additional considerations need to be accounted for, such as contrast agent relaxivity, baseline R_1 ($1/T_1$), change in R_1 due to contrast agent, and the robustness of arterial input function measurement, all of which present specific requirements for measurement and QC [83]. The analysis of DCE data requires the choice of a pharmacokinetic model that is appropriate for the data [84,85] and, as with many qMR parametric modelling approaches, these DCE models all have different assumptions and flaws. These choices all have significant consequences on the summary statistic that is often the endpoint of the analysis and data inputted into a clinical trial case report form. Although some attempts have been made [86], a well-characterised traceable physical phantom is not available for defining the accuracy or repeatability of K^{trans}, meaning these must be assessed in software phantoms [87] and a relevant clinical population. Many publications have shown K^{trans} to change following treatment [80,88], but absolute values are not well reported and have poor reproducibility between centres, likely due to differences in acquisition and analysis. A recent multi-centre study designed and tested a QA framework based on phantom measurements for T_2, T_1, DWI and DCE-MRI for cervical cancer. They demonstrated consistent T_2 and ADC values across sites, but found that DCE metrics and T_1 mapping to be more challenging [89].

.



Box 1 Case study of the validation of ADC in oncology: (a) accuracy of ADC using standardised phantoms; (b) physics consideration arising from specific imperfections in MR engineering; (c) repeatability of clinical ADC measurements and; (d) reproducibility across different scanner and vendors. **A1 & A2** – Reprinted from Chenevert *et al.* Journal of Magnetic Resonance Imaging. 2011;34(4):983–7. [90], with permission from John Wiley and Sons. Copyright © 2010 Wiley-Liss, Inc. **A3 & A4** - Republished with permission from American Journal of Neuroradiology, from Palacios *et al.* American Journal of Neuroradiology. 2017;38(3):537–45, [91]. Copyright © 2017 by American Journal of Neuroradiology. **B1** - Reprinted from Malyarenko *et al.* Magnetic Resonance in Medicine. 2014;71(3):1312–23 [92], with permission from John Wiley and Sons. Copyright © 2013 Wiley Periodicals, Inc. **B2** - Reprinted from Malyarenko *et al.* Journal of Magnetic Resonance Imaging. 2013;37(5):1238–46. [93], with permission from John Wiley and Sons. Copyright © 2012 Wiley Periodicals, Inc. **B3** – [94]. **C1** – Reprinted from Winfield *et al.* Radiology. 2017;284(1):88–99 [95], with permission from Radiological Society of North America, Copyright © 2017. Radiological Society of North America. **C2** – Reprinted by permission from Springer Nature: Weller *et al.* European radiology. 2017;27(11):4552–62. [96]. Copyright © 2017, European Society of Radiology. **C3** - Reprinted from Spick *et al.* NMR in Biomedicine. 2016;29(10):1445–53. [97], with permission from John Wiley and Sons. Copyright © 2016 John Wiley & Sons, ltd. **D1** – Reprinted by permission from Springer Nature: Michoux *et al.* European Radiology. 2021: 31:4514–27. [98]. Copyright © 2021, European Society of Radiology. **D2** - Reprinted from Donati *et al.* Radiology. 2014;270(2):454–63. [99], with permission from Radiological Society of North America. Copyright © 2014, Radiological Society of North America. **D3** – [2,14].

Some qMR IBs are long-standing and have translated into clinically-used community IBs by virtue of having obvious clinical utility, such as the evaluation of tumour size and spread using TNM (tumour, node, and metastasis) staging or estimates of left ventricular ejection fraction in cardiac MR [18]. Others have moved into the clinic via a commercial route, such as the use of Ferriscan as a companion diagnostic to measure liver iron concentration using patented technology based on R_2 measurements [51] and multi-parametric liver imaging using relaxometry and fat fraction methods [52–55]. Interestingly, both fat fraction and T_2-based methods show the highest translation ratio between journal articles and clinical database studies in Fig. 2c, supporting the idea that commercialisation assists translation. Although commercial qMR IBs are often not 'state-of-the-art', they benefit from being better understood, with known failure modes, repeatability, reproducibility and use cases. This commercial pragmatism is a useful feature of industry and helps methods gain real clinical utility.

To date, only one qMR IB has been approved by the FDA via the Biomarker Qualification Program [56]. This is a prognostic biomarker with patient age and baseline glomerular filtration rate for autosomal dominant polycystic kidney disease using total kidney volume (TKV) as assessed by MRI, CT and US. Another six MR qIBs have been submitted, these are: 1) a prognostic anatomical biomarker to identify patients more likely to experience knee osteoarthritis disease progression; 2) a pharmacodynamic/response anatomic biomarker for Crohn's disease used as a co-primary endpoint; 3) a diagnostic enrichment biomarker for selecting patients for non-alcoholic steatohepatitis (NASH) trials using the proton density fat fraction (PDFF) of liver; 4) a safety biomarker indicating potential intrahepatic drug-drug interactions using the liver-specific contrast agent, gadoxetate; 5) a prognostic enrichment biomarker to identify NASH patients more likely to experience clinical endpoints such as progression to cirrhosis, hepatic decompensation events during the timeframe of a NASH clinical trial using corrected T_1 (cT_1) as assessed by MRI and serum biomarkers as assessed by enhanced liver fibrosis test; and 6) a diagnostic enrichment biomarker intended for use, in conjunction with clinical factors, to identify patients likely to have liver histopathologic findings of NASH and with a non-alcoholic fatty liver disease using PDFF, cT1 and 2D and 3D MRE.

Other routes to translation are qualification of novel methodologies via the European Medicine Agency (EMA) [100] and, within the UK, recommended use via the National Institute for Health and Care Excellence (NICE). NICE have a number of programmes in which imaging-related technologies may be evaluated: the Medical Technologies Evaluation Programme (MTEP) [101], the Diagnostics Assessment Programme (DAP) [102], and the Technology Appraisal Guidance (TAG) [103]. DAP is suitable for evaluating diagnostic tests and technologies where such evaluation is complex and that have the potential to improve health outcomes, but whose introduction is likely to be associated with an overall increase in cost to the NHS. Imaging technologies that may offer similar health outcomes at less cost, or improved health outcomes at the same cost as current NHS or which are associated with therapeutics are likely to be more suitable for evaluation by the MTEP. A companion diagnostic technology, where the primary purpose of the technology is to identify patients who respond best to new drugs, may be suitable for evaluation under TAG in the context of an appraisal of the drug to which it is linked. A new companion diagnostic for established drugs, may be more suitable for assessment by DAP. In all cases, these programmes anticipate commercial involvement by requiring a 'Product Sponsor', who is likely the manufacturer, developer, distributor or agent of a relevant technology.

There is a largely untapped strategic advantage in the UK from a translational perspective – there is an integrated healthcare system with clinicians and researchers working closely within the NHS framework, with access to a variety of scanner vendors over many centres acquiring data that, if utilised fully, could allow ready translation of qMR IBs into clinical practice. The development of multicentre studies and standardised protocols within the emerging NHS Regional Imaging Networks [22,104] and the developing support of the NCITA MR Core Lab [105] has the potential to augment this advantage.

6. Gap analysis of qMR validation and quality management

Based on the evidence collected from the literature and clinical research study database searches and the QA/QC survey we have been able to report an overview of current practice in qMR use and quality management. To perform a gap analysis, we must provide a standard from which to assess the gap. That standard is established in metrology [60,106,107]. Box 2 gives an overview of metrology in relation to qMR IBs and Table 2 provides a gap analysis of relevant quality management domains and encapsulates the findings of this study with suggested best practice, the gaps and the challenges we face to bridge those gaps.

Box 2 An overview of metrology in relation to qMR imaging biomarkers.

Metrology and qMR imaging biomarkers.

* Quantitative imaging is a form of measurement. The quantity of interest and, more importantly, our knowledge of it varies depending on how the image is formed. Any quantitatively estimated value of the quantity is the result of a measurement. **The quantity intended to be measured is known as a measurand.** MRI-based measurands include relaxation constants such as T_1 and T_2, and also measures of size, such as cortical or ventricular volume. Metrology as a field seeks to study measurements of all kinds and has some useful unifying principles which can be applied in any measurement context, including MRI.
* It is common to talk about measurements in terms of their accuracy and precision. **Precision refers to the spread of a set of values from similar measurements of the same quantity, and accuracy refers to the closeness of agreement of a measured value from a "true" value of the quantity.** These terms are useful, but both implicitly assume the existence of some perfect, underlying true quantity value, sometimes referred to as a "ground truth". The idea of a ground truth is intuitively appealing, it is not necessarily helpful in characterising a measurement.
* In practice a measurement can be compared to one made by an alternate method which is known to be more accurate and precise (but perhaps is less practical or too expensive), but this measurement is not a "ground truth". That measurement in turn can be assessed by comparing it to another non-perfect measurement**. It is impossible, even in principle, for any measurement to be infinitely precise and completely accurate and hence eventually this chain of comparisons reaches a limit.** As such, the "accuracy" of a measurement can never be truly known – the concept is not fully quantitative.
* There are concepts which are fully quantitative and generic in metrology, however. **Every measurement is made up of two elements: a numerical value and a unit, which acts as a reference to compare the measurement against.** As such, we need firstly to know a numerical value, and the how (and how well) the measurement is calibrated to the reference unit.
* The limitations of a measurement are captured in its uncertainty. Uncertainty expresses the limits of our knowledge of the quantity and is related to how the measurement is made. Uncertainty includes systematic effects, which introduce a consistent bias to measurements and random effects, which are stochastic but quantifiable. **A typical measurement uncertainty will be quoted as a confidence interval, which gives a probability that the (unknown) "true value" is in a certain range.** Note that it is distinct from the related idea of an error, which is the difference between the measured and true values. Since the true value is unknown, the error can never be known. Unlike error, uncertainty is quantifiable.
* **We also need to know how well the measurement is calibrated to the unit in question by comparing our measurement to other reference measurements.** A particular measurement can be made of a standard object (a local standard), which is itself calibrated to another held at an accreditation laboratory (a secondary standard). The accreditation laboratory's standard is then calibrated to a primary standard held by a National Measurement Institute (NMI) such as National Physical Laboratory (NPL) or National Institute of Standards and Technology (NIST) which is in turn calibrated to the definition of the units in question. This is the concept of traceability.
* **Via this chain of comparisons, and the associated calibration certificates, it is possible to trace (almost) any measurement back to a common set of agreed standards**, allowing consistency and comparability between measurements made in different ways, in different places, using different procedures and equipment.
* Notice that the focus here is not what is being measured, but how well it can be measured. **In strict metrological terms, a measurement result consists of the value measured, and an associated uncertainty (range of values) and a confidence interval that measurement repeats would fall in the specified range.** Once we have this the notion of a perfect ground truth becomes unnecessary.
* **Measurements themselves can be analysed using an uncertainty budget.** Here, each step of the measurement is broken out and characterised by its own

(*continued on next page*)

(*continued*)

uncertainty, which are frequently simpler to estimate. The contribution of the uncertainty from each aspect of the measurement can then be combined into a single value for the entire process and allow a measurement to be characterised without knowing a ground truth.

* **Similarly, aspects of a measurement may themselves be made traceable**. A good example of this is the timing reference used in MR scanners: pulse timing can be calibrated to a national or international timing reference so provide a reliable estimate of the uncertainty in pulse timings. Similar approaches can be applied to field strengths and gradients (via frequency).

* Quantitative MRI is currently at a transition – we know how to make quantitative measurements, but even though the components of our scanners are well-calibrated there is not yet a metrological system in place to provide traceability to SI units and primary standards for the measurements we make from images. **With an appropriate system in place to calibrate and benchmark scanners and acquisitions, the differences between measurements from different systems becomes comprehensible and quantitative comparisons become easier and more reliable.**

7. Conclusions

The quantitative use of MRI biomarkers has the potential to transform clinical decision making. This work presents an overview of the current use of qMR IBs in the UK and the associated quality management practices. It highlights the trend towards an increasing number of journal articles, clinical trials and multicentre studies using qMR IBs. This is true of all qMR methods, particularly popular methods such as DWI, DTI and volumetry. However, the "translation ratio" of journal articles in 2011–2015 to studies recorded in clinical research databases in 2016–2020 was found to be much lower for these popular academic techniques than for qMR methods that have translated into clinical use via a commercial route, such as fat fraction and T_2 mapping.

Although the number of journal articles citing quality management terms doubled between the periods 2011–2015 and 2016–2020, the proportion of all journal articles only increased by 3.0 % and the survey suggested that basic QA procedures are underreported in the literature. The survey also revealed that the access to quantitative phantoms for QA was reasonably good and that the quality management of data acquisition was more comprehensive than the QC of the acquired data and data analysis procedures. As such, the standardisation and reporting of the quality management aspects of a study and moves towards transparency and sharing of imaging data and analysis code must be more strongly encouraged.

The standardisation and subsequent translation of qMR IBs require improved quality management of the entire pipeline from study set-up, through data acquisition and analysis, to publication. The gap analysis of current practice to the metrological ideal gives us a platform from which to assess how best to traverse the gaps and improve clinical translation of qMR IBs.

This manuscript is designed to provide an overview of the current status of qMR use and quality management from a UK perspective, but with a view of using it as a platform to develop worldwide community consensus on the best ways in which we can support each other to enhance current progress, improve current shortfalls and provide robust, validated, metrologically-valid quantitative MR imaging biomarkers to improve patient care in the years ahead.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

PC, JOC, EO, JMW, SK and MJO gratefully acknowledge financial support from the Cancer Research UK National Cancer Imaging Translational Accelerator Award (C4278/A27066).

JMW and SK acknowledge funding from Cancer Research UK and Engineering and Physical Sciences Research Council support to the Cancer Imaging Centre at the Institute of Cancer Research and Royal Marsden Hospital in association with the Medical Research Council and Department of Health C1060/A10334, C1060/A16464 and National Health Service funding to the National Institute of Health and Care Research Biomedical Research Centre, Experimental Cancer Medicine Centre, the Clinical Research Facility in Imaging, and the Cancer Research Network. The views expressed in this publication are those of the author(s) and not necessarily those of the National Health Service, the National Institute of Health and Care Research or the Department of Health.

EO acknowledges funding from the Imperial College NIHR Biomedical Research Centre award (WSCC_P62585) and the Medical Research Council grant (MC-A652-5PY80)

GP acknowledges support from the Engineering and Physical Sciences Research Council (EPSRC) award EP/S031510/1.

We also thank the survey responders and Prof. Thomas Chenevert, Prof. Pratik Mukherjee, Prof. Martin Leach, Dr Alex Weller, Prof. Pascal Baltzer, Dr Nicolas Michoux, Prof. Harriet Thoeny and Dr Christina Messiou, for figures reproduced in this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2022.08.015.

References

[1] Gulani V, Seiberlich N. Quantitative MRI: Rationale and Challenges 2020: xxxvii–li. https://doi.org/10.1016/B978-0-12-817057-1.00001-9.

[2] Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. J Magn Reson Imaging 2019;49: e101–21. https://doi.org/10.1002/JMRI.26518.

[3] deSouza NM, Achten E, Alberich-Bayarri A, Bamberg F, Boellaard R, Clément O, et al. Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL* subcommittee of the European Society of Radiology (ESR). Insights. Imaging 2019;10(1). https://doi.org/10.1186/s13244-019-0764-0.

[4] Food and Drug Administration. Clinical Trial Imaging Endpoint Process Standards Guidance for Industry Clinical/Medical Clinical Trial Imaging Endpoint Process Standards Guidance for Industry 2018.

[5] Keenan KE, Biller JR, Delfino JG, Boss MA, Does MD, Evelhoch JL, et al. Recommendations towards standards for quantitative MRI (qMRI) and Outstanding Needs HHS Public Access. J Magn Reson Imaging 2019;49(7): e26–39.

[6] Deng J, Wang Y. Quantitative magnetic resonance imaging biomarkers in oncological clinical trials: Current techniques and standardization challenges. Chronic Dis Transl Med 2017;3:8–20. https://doi.org/10.1016/j.cdtm.2017.02.002.

[7] McGee KP, Hwang K, Sullivan DC, Kurhanewicz J, Hu Y, Wang J, et al. Magnetic resonance biomarkers in radiation oncology: The report of AAPM Task Group 294. Med Phys 2021;48(7). https://doi.org/10.1002/MP.14884.

[8] Keenan KE, Gimbutas Z, Dienstfrey A, Stupic KF, Boss MA, Russek SE, et al. Multi-site, multi-platform comparison of MRI T1 measurement using the system phantom. PLoS ONE 2021;16(6):e0252966. https://doi.org/10.1371/journal.pone.0252966.

[9] 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: Quantitative MRI Phantoms Review. Magn Reson Med 2018;79 (1):48–61.

[10] Mazzoni LN, Bock M, Levesque IR, Lurie DJ, Palma G. New developments in MRI: System characterization, technical advances and radiotherapy applications. Phys Med 2021;90:50–2. https://doi.org/10.1016/j.ejmp.2021.09.001.

[11] ISO 9000: Quality management systems — Fundamentals and vocabulary https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en (accessed February 1, 2022).

[12] Farahani K, Tata D, Nordstrom R. QIN benchmarks for clinical translation of quantitative imaging tools. Benchmarks for Clinical Translation of Quantitative Imaging Tools 2019;5(1):1–6.

[13] Press RH, Shu H-K, Shim H, Mountz JM, Kurland BF, Wahl RL, et al. The use of quantitative imaging in radiation oncology: A Quantitative Imaging Network (QIN) perspective radiation oncology. Int J Radiat Oncol Biol Phys 2018;102(4): 1219–35.

[14] QIBA profiles http://qibawiki.rsna.org/index.php/Profiles (accessed February 1, 2022).
[15] Belli G, Busoni S, Ciccarone A, Coniglio A, Esposito M, Giannelli M, et al. Quality assurance multicenter comparison of different MR scanners for quantitative diffusion-weighted imaging. J Magn Reson Imaging 2016;43(1):213–9.
[16] Fedeli L, Belli G, Ciccarone A, Coniglio A, Esposito M, Giannelli M, et al. Dependence of apparent diffusion coefficient measurement on diffusion gradient direction and spatial position – A quality assurance intercomparison study of forty-four scanners for quantitative diffusion-weighted imaging. Phys Med 2018; 55:135–41.
[17] Fedeli L, Benelli M, Busoni S, Belli G, Ciccarone A, Coniglio A, et al. On the dependence of quantitative diffusion-weighted imaging on scanner system characteristics and acquisition parameters: A large multicenter and multiparametric phantom study with unsupervised clustering analysis. Phys Med 2021;85:98–106.
[18] O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14(3): 169–86.
[19] National Library of Medicine. 2020 MeSH Pubtypes https://www.nlm.nih.gov/mesh/pubtypes.html (accessed July 14, 2021).
[20] Shea DA, Morgan D. CRS report for congress the helium-3 shortage. Supply, demand, and options for Congress 2010.
[21] NHS. The NHS Long Term Plan 2019. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf (accessed February 1, 2022).
[22] NHS. Diagnostic imaging network implementation guide 2021. https://www.england.nhs.uk/wp-content/uploads/2021/04/B0030-Implementation-guide.pdf (accessed February 1, 2022).
[23] NHS. Diagnostic imaging network capital equipment planning guide 2021. https://www.england.nhs.uk/wp-content/uploads/2021/04/B0030-Capital-equipment-planning-guide-April-2021.pdf (accessed February 1, 2022).
[24] Delis H, Christaki K, Healy B, Loreti G, Poli GL, Toroi P, et al. Moving beyond quality control in diagnostic radiology and the role of the clinically qualified medical physicist. Phys Med 2017;41:104–8.
[25] Stikov N, Trzasko JD, Bernstein MA. Reproducibility and the future of MRI research. Magn Reson Med 2019;82:1981–3. https://doi.org/10.1002/mrm.27939.
[26] MR Together https://mritogether.github.io/ (accessed February 1, 2022).
[27] Prostate cancer: diagnosis and management NICE guideline 2019. www.nice.org.uk/guidance/ng131 (accessed January 24, 2022).
[28] Alsop DC, Detre JA, Golay X, Günther M, Hendrikse J, Hernandez-Garcia L, et al. Recommended implementation of arterial spin-labeled Perfusion mri for clinical applications: A consensus of the ISMRM Perfusion Study group and the European consortium for ASL in dementia. Magn Reson Med 2015;73(1):102–16.
[29] deSouza NM, Winfield JM, Waterton JC, Weller A, Papoutsaki M-V, Doran SJ, et al. Implementing diffusion-weighted MRI for body imaging in prospective multicentre trials: current considerations and future perspectives. Eur Radiol 2018;28(3):1118–31.
[30] Barnes A, Alonzi R, Blackledge M, Charles-Edwards G, Collins DJ, Cook G, et al. UK quantitative WB-dWI technical workgroup: consensus meeting recommendations on optimisation, quality control, processing and analysis of quantitative whole-body diffusion-weighted imaging for cancer. Br J Radiol 2018; 91(1081):20170577.
[31] Boxerman JL, Quarles CC, Hu LS, Erickson BJ, Gerstner ER, Smits M, et al. Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in highgrade gliomas. Neuro Oncol 2020;22(9):1262–75.
[32] Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, et al. Diffusion-weighted imaging of the breast-a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging working group. Eur Radiol 2020;30(3):1436–50.
[33] Padhani AR, Liu G, Mu-Koh D, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 2009;11(2):102–25.
[34] Dyverfeldt P, Bissell M, Barker AJ, Bolger AF, Carlhäll C-J, Ebbers T, et al. 4D cardiovascular magnetic resonance consensus statement. J Cardiovasc Magn Reson 2015;17:72. https://doi.org/10.1186/s12968-015-0174-5.
[35] Rovira À, Wattjes MP, Tintoré M, Tur C, Yousry TA, Sormani MP, et al. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. Nat Rev | Neurol 2015;11(8):471–82.
[36] Bernasconi A, Cendes F, Theodore WH, Gill RS, Koepp MJ, Hogan RE, et al. Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: A consensus report from the international league against epilepsy neuroimaging task force. Epilepsia 2019. https://doi.org/10.1111/epi.15612.
[37] Ljimani A, Caroli A, Laustsen C, Francis S, Mendichovszky IA, Bane O, et al. Consensus-based technical recommendations for clinical translation of renal diffusion-weighted MRI. Magn Reson Mater Physics, Biol Med 2020;33(1): 177–95.
[38] Nery F, Buchanan CE, Harteveld AA, Odudu A, Bane O, Cox EF, et al. Consensus-based technical recommendations for clinical translation of renal ASL MRI. Magn Reson Mater Physics, Biol Med 2020;33(1):141–61.
[39] Mendichovszky I, Pullens P, Dekkers I, Nery F, Bane O, Pohlmann A, et al. Technical recommendations for clinical translation of renal MRI: a consensus project of the Cooperation in Science and Technology Action PARENCHIMA. Magn Reson Mater Physics, Biol Med 2020;33(1):131–40.
[40] Dekkers IA, de Boer A, Sharma K, Cox EF, Lamb HJ, Buckley DL, et al. Consensus-based technical recommendations for clinical translation of renal T1 and T2 mapping MRI. Biol Med 2020;33(1):163–76.
[41] 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(1):48–61.
[42] Zhang Q, Werys K, Popescu IA, Biasiolli L, Ntusi NAB, Desai M, et al. Quality assurance of quantitative cardiac T1-mapping in multicenter clinical trials - A T1 phantom program from the hypertrophic cardiomyopathy registry (HCMR) study. Int J Cardiol 2021;330:251–8.
[43] Karakuzu A, Biswas L, Cohen-Adad J, Stikov N. Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. BioRxiv 2022;1:2021.12.27.474259.
[44] Software Sustainability Institute https://www.software.ac.uk/ (accessed February 18, 2022).
[45] Github https://github.com/ (accessed February 18, 2022).
[46] Gitlab https://gitlab.com/ (accessed February 18, 2022).
[47] Docker Hub https://hub.docker.com/ (accessed February 23, 2022).
[48] Singularity https://sylabs.io/singularity (accessed February 27, 2022).
[49] Open Science Initiative for Perfusion Imaging https://www.osipi.org/ (accessed February 10, 2022).
[50] ISMRM MR-Hub https://ismrm.github.io/mrhub/ (accessed February 10, 2022).
[51] Ferriscan - Resonance Health https://www.resonancehealth.com/products/ferriscan-mri-measurement-of-liver-iron-concentration.html (accessed February 18, 2022).
[52] LiverMultiScan - Perspectum https://perspectum.com/products/livermultiscan (accessed February 18, 2022).
[53] IDEAL IQ - GE Healthcare https://www.gehealthcare.co.uk/en/products/magnetic-resonance-imaging/mr-applications/ideal-iq—body (accessed February 18, 2022).
[54] mDIXON Quant - Philips https://www.philips.co.uk/healthcare/product/HCNMRB462/mdixon-quant-mr-clinical-application (accessed February 18, 2022).
[55] LiverLab - Siemens Healthineers https://www.siemens-healthineers.com/en-uk/magnetic-resonance-imaging/options-and-upgrades/clinical-applications/liver-lab (accessed February 18, 2022).
[56] FDA Biomarker Qualification Program https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program (accessed February 18, 2022).
[57] Yankeelov TE. The Quantitative Imaging Network: A decade of achievement. Tomography 2019;5:A8. https://doi.org/10.18383/j.tom.2019.00999.
[58] ECOG-ACRIN Imaging Core Lab https://ecog-acrin.org/research-cores/imaging-core-laboratory/ (accessed February 1, 2022).
[59] EIBALL https://www.myesr.org/research/european-imaging-biomarkers-alliance-eiball (accessed February 18, 2022).
[60] Alberich-Bayarri A, Sourbron S, Golay X, deSouza N, Smits M, van der Lugt A, et al. ESR Statement on the Validation of Imaging Biomarkers. Eur Soc Radiol 2020;11(1). https://doi.org/10.1186/s13244-020-00872-9.
[61] EIBALL Biomarker Inventory https://www.myesr.org/research/biomarkers-inventory (accessed February 23, 2022).
[62] Weingärtner S, Desmond KL, Obuchowski NA, Baessler B, Zhang Y, Biondetti E, et al. Development, validation, qualification, and dissemination of quantitative MR methods: Overview and recommendations by the ISMRM quantitative MR study group. Magn Reson Med 2022;87(3):1184–206.
[63] Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. J Magn Reson Imaging 2006;23:827–39. https://doi.org/10.1002/JMRI.20583.
[64] Sghedoni R, Coniglio A, Mazzoni LN, Busoni S, Belli G, Tarducci R, et al. A straightforward multiparametric quality control protocol for proton magnetic resonance spectroscopy: Validation and comparison of various 1.5 T and 3 T clinical scanner systems. Phys Med 2018;54:49–55.
[65] The Osteoarthritis Initiative (OAI) https://nda.nih.gov/oai/ (accessed February 18, 2022).
[66] Alzheimer's Disease Neuroimaging Initiative (ADNI) http://adni.loni.usc.edu/ (accessed February 18, 2022).
[67] IMI/IHI projects https://www.imi.europa.eu/about-imi/innovative-health-initiative (accessed February 18, 2022).
[68] McAteer MA, O'Connor JPB, Koh DM, Leung HY, Doran SJ, Jauregui-Osoro M, et al. Introduction to the National Cancer Imaging Translational Accelerator (NCITA): a UK-wide infrastructure for multicentre clinical translation of cancer imaging biomarkers. Br J Cancer 2021;125(11):1462–5.
[69] Koh DM, Collins DJ. Diffusion-weighted MRI in the body: Applications and challenges in oncology. Am J Roentgenol 2007;188:1622–35. https://doi.org/10.2214/AJR.06.1403.
[70] Charles-Edwards EM. Diffusion-weighted magnetic resonance imaging and its application to cancer. Cancer Imaging 2006;6:135–43. https://doi.org/10.1102/1470-7330.2006.0021.
[71] Sinkus R, Van Beers BE, Vilgrain V, Desouza N, Waterton JC. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. Eur J Cancer 2012;48:425–31. https://doi.org/10.1016/j.ejca.2011.11.034.
[72] Winfield JM, Blackledge MD, Tunariu N, Koh DM, Messiou C. Whole-body MRI: a practical guide for imaging patients with malignant bone disease. Clin Radiol 2021;76:715–27. https://doi.org/10.1016/j.crad.2021.04.001.
[73] Lee MK, Choi Y, Jung S-L. Diffusion-weighted MRI for predicting treatment response in patients with nasopharyngeal carcinoma: a systematic review and meta-analysis. Sci Rep 2021;11(1).

[74] Nalaini F, Shahbazi F, Mousavinezhad SM, Ansari A, Salehi M. Diagnostic accuracy of apparent diffusion coefficient (ADC) value in differentiating malignant from benign solid liver lesions: a systematic review and meta-analysis. Br J Radiol 2021;94(1123):20210059.

[75] Wielema M, Dorrius MD, Pijnappel RM, De Bock GH, Baltzer PAT, Oudkerk M, et al. Diagnostic performance of breast tumor tissue selection in diffusion weighted imaging: A systematic review and meta-analysis. Diagnostic performance of breast tumor tissue selection in diffusion weighted imaging: A systematic review and meta-analysis 2020;15(5):e0232856.

[76] Alexander DC, Dyrby TB, Nilsson M, Zhang H. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR Biomed 2019;32:e3841.

[77] Slator PJ, Palombo M, Miller KL, Westin C, Laun F, Kim D, et al. Combined diffusion-relaxometry microstructure imaging: Current status and future prospects. Magn Reson Med 2021;86(6):2987–3011.

[78] Afzali M, Pieciak T, Newman S, Garyfallidis E, Özarslan E, Cheng H, et al. The sensitivity of diffusion MRI to microstructural properties and experimental factors. J Neurosci Methods 2021;347:108951.

[79] Nilsson M, Van Westen D, Ståhlberg F, Sundgren PC, Lätt J, Nilsson M, et al. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. Magn Reson Mater Phy 2013;26:345–70. https://doi.org/10.1007/s10334-013-0371-x.

[80] O'Connor JPB, Jackson A, Parker GJ, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. Nat Rev Clin Oncol 2012;9:167–77. https://doi.org/10.1038/nrclinonc.2012.2.

[81] Du R, Cai J, Cui B, Wu H, Zhao XQ, Ye P. Rapid improvement in carotid adventitial angiogenesis and plaque neovascularization after rosuvastatin therapy in statin treatment–naïve subjects. J Clin Lipidol 2019;13:847–53. https://doi.org/10.1016/j.jacl.2019.07.008.

[82] Waterton JC, Ho M, Nordenmark LH, Jenkins M, DiCarlo J, Guillard G, et al. Repeatability and response to therapy of dynamic contrast-enhanced magnetic resonance imaging biomarkers in rheumatoid arthritis in a large multicentre trial setting. Eur Radiol 2017;27(9):3662–8.

[83] Kim H. Variability in Quantitative DCE-MRI: Sources and Solutions. J Nat Sci 2018;4:1–16.

[84] Sourbron SP, Buckley DL. Tracer kinetic modelling in MRI: Estimating perfusion and capillary permeability. Phys Med Biol 2012;57(2):R1–33.

[85] Khalifa F, Soliman A, El-Baz A, Abou El-Ghar M, El-Diasty T, Gimel'farb G, et al. Models and methods for analyzing DCE-MRI: A review. Med Phys 2014;41(12):124301.

[86] Anderson JR, Ackerman JJH, Garbow JR. Semipermeable Hollow Fiber Phantoms for Development and Validation of Perfusion-Sensitive MR Methods and Signal Models. Concepts Magn Reson Part B Magn Reson Eng 2011;39B:149–58. https://doi.org/10.1002/cmr.b.

[87] Barboriak Lab; Duke Univeristy. QIBA digital reference object https://sites.duke.edu/dblab/qibacontent/.

[88] Little RA, Barjat H, Hare JI, Jenner M, Watson Y, Cheung S, et al. Evaluation of dynamic contrast-enhanced MRI biomarkers for stratified cancer medicine: How do permeability and perfusion vary between human tumours? Magn Reson Imaging 2018;46:98–105.

[89] van Houdt PJ, Kallehauge JF, Tanderup K, Nout R, Zaletelj M, Tadic T, et al. Phantom-based quality assurance for multicenter quantitative MRI in locally advanced cervical cancer. Radiother Oncol 2020;153:114–21.

[90] Chenevert TL, Galbán CJ, Ivancevic MK, Rohrer SE, Londy FJ, Kwee TC, et al. Diffusion coefficient measurement using a temperature controlled fluid for quality control in multi-center studies. J Magn Reson Imaging 2011;34(4):983–7.

[91] Palacios EM, Martin AJ, Boss MA, Ezekiel F, Chang YS, Yuh EL, et al. Toward Precision and Reproducibility of Diffusion Tensor Imaging: A Multicenter Diffusion Phantom and Traveling Volunteer Study. AJNR Am J Neuroradiol 2017;38(3):537–45.

[92] Malyarenko DI, Ross BD, Chenevert TL. Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements. Magn Reson Med 2014;71:1312–23. https://doi.org/10.1002/mrm.24773.

[93] Malyarenko D, Galbán CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, et al. Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging 2013;37(5):1238–46.

[94] Newitt DC, Tan ET, Wilmes LJ, Chenevert TL, Kornak J, Marinelli L, et al. Gradient nonlinearity correction to improve apparent diffusion coefficient accuracy and standardization in the American college of radiology imaging network 6698 breast cancer trial. J Magn Reson Imaging 2015;42(4):908–19.

[95] Winfield JM, Tunariu N, Rata M, Miyazaki K, Jerome NP, Germuska M, et al. Extracranial soft-tissue tumors: repeatability of apparent diffusion coefficient estimates from diffusion-weighted MR imaging. Radiology 2017;284(1):88–99.

[96] Weller A, Papoutsaki MV, Waterton JC, Chiti A, Stroobants S, Kuijer J, et al. Diffusion-weighted (DW) MRI in Lung Cancers: ADC Test-retest Repeatability Europe PMC Funders Group. Eur Radiol 2017;27(11):4552–62.

[97] Spick C, Bickel H, Pinker K, Bernathova M, Kapetas P, Woitek R, et al. Diffusion-weighted MRI of breast lesions: a prospective clinical investigation of the quantitative imaging biomarker characteristics of reproducibility, repeatability, and diagnostic accuracy. NMR Biomed 2016;29(10):1445–53.

[98] Michoux NF, Ceranka JW, Vandemeulebroucke J, Peeters F, Lu P, Absil J, et al. Repeatability and reproducibility of ADC measurements: a prospective multicenter whole-body-MRI study. Eur Radiol 2021;31(7):4514–27.

[99] Donati OF, Chong D, Nanz D, Boss A, Froehlich JM, Andres E, et al. Diffusion-weighted MR imaging of upper abdominal organs: Field strength and intervendor variability of apparent diffusion coefficients. Radiology 2014;270(2):454–63.

[100] European Medicine Agency Qualification https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/qualification-novel-methodologies-medicine-development-0 (accessed February 1, 2022).

[101] NHS NICE. Medical Technologies Evaluation Programme https://www.nice.org.uk/process/pmg33/resources/medical-technologies-evaluation-programme-methods-guide-pdf-72286774205893 (accessed February 10, 2022).

[102] NHS NICE. Diagnostic Assessment Programme https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf (accessed February 10, 2022).

[103] NHS NICE. Technology Appraisal Guidance https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance.

[104] NHS. Transforming imaging services in England: a national strategy for imaging networks 2019:55. https://www.england.nhs.uk/transforming-imaging-services-in-england/ (accessed February 1, 2022).

[105] NCITA MR Core Lab ncita.org.uk/mr-core-lab.

[106] 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(1120):20201215.

[107] Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015;277(3):813–25.

[108] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Chan A-W, et al. Consensus statement Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020;26(9):1364–74.