Development, Validation, Qualification and Dissemination of Quantitative MR Methods: Overview and Recommendations by the ISMRM Quantitative MR Study Group

Journal:	*Magnetic Resonance in Medicine*
Manuscript ID	MRM-21-22505.R1
Wiley - Manuscript type:	Guidelines
Research Type:	Technique Development
Research Focus:	No specific tissue or organ focus

Development, Validation, Qualification and Dissemination of Quantitative MR Methods: Overview and Recommendations by the ISMRM Quantitative MR Study Group

Sebastian Weingärtner, PhD*[,1]; Kimberly L. Desmond, PhD*[2,3]; Nancy Obuchowski, PhD[4]; Bettina Baessler, MD[5]; Yuxin Zhang, PhD[6,7]; Emma Biondetti, PhD[8]; Dan Ma, PhD[9]; Xavier Golay, PhD[10]; Michael A. Boss, PhD[11]; Jeffrey L. Gunter, PhD[12]; Kathryn E. Keenan, PhD[13]; Diego Hernando, PhD[6,7]; on behalf of the ISMRM Quantitative MR Study Group

1. Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands
2. Brain Health Imaging Centre, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
3. Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada
4. Department of Biostatistics, Cleveland Clinic, Cleveland, OH, USA
5. Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Zurich, Switzerland
6. Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA
7. Department of Radiology, University of Wisconsin-Madison, Madison, WI, USA
8. Department of Neurosciences, Imaging and Clinical Sciences, Gabriele d'Annunzio University of Chieti and Pescara, Chieti, Italy
9. Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA
10. Brain Repair & Rehabilitation, Institute of Neurology, University College London, United Kingdom
11. American College of Radiology, Philadelphia, PA, USA
12. Department of Radiology, Mayo Clinic, Rochester, MN, USA
13. National Institute of Standards and Technology, Boulder, CO, USA

Word count: 7500

Corresponding author:

Commented [A1]: R2.1

Commented [A2]: E.4

Diego Hernando, PhD
Departments of Radiology and Medical Physics
University of Wisconsin-Madison, Madison, 53705, USA
Email: dhernando@wisc.edu
Twitter: @dherarr

* Sebastian Weingärtner and Kimberly L. Desmond contributed equally to this work

Abstract:

On behalf of the International Society for Magnetic Resonance in Medicine (ISMRM) Quantitative MR Study Group, this article provides an overview of considerations for the development, validation, qualification, and dissemination of quantitative MR (qMR) methods. This process is framed in terms of two central technical performance properties, i.e., bias and precision. Although qMR is confounded by undesired effects, methods with low bias and high precision can be iteratively developed and validated. For illustration, two distinct qMR methods are discussed throughout the manuscript: quantification of liver proton-density fat fraction, and cardiac T_1. These examples demonstrate the expansion of qMR methods from research centers toward widespread clinical dissemination. The overall goal of this article is to provide trainees, researchers and clinicians with essential guidelines for the development and validation of qMR methods, as well as an understanding of necessary steps and potential pitfalls for the dissemination of quantitative MR in research and in the clinic.

Keywords: Quantitative, bias, precision, confounding factors, PDFF, T_1

Commented [A3]: R2.2

Commented [A4]: R1.1

INTRODUCTION

Magnetic resonance probes a wide array of tissue contrasts, spectral properties and anatomical information. Based on this wealth of contrast mechanisms, a variety of quantitative MR (qMR) methods that extract quantifiable information from MR acquisitions (1-3) have been proposed and continue to emerge from the MR research community. Upon successful development and validation, qMR methods enable improved standardization in the detection, staging, and treatment monitoring of diseases, both in research and in clinical practice (4-7). On behalf of the International Society for Magnetic Resonance in Medicine (ISMRM) Quantitative MR Study Group, we provide an overview of the process of development of qMR methods, as well as guidelines for their technical validation, clinical qualification, application, and dissemination. To illustrate this process, we provide examples from two distinct qMR methods: quantification of liver proton-density fat fraction (PDFF), and T_1 quantification in the myocardium (cardiac T_1 mapping; see Figure 1). These two methods were selected based on their substantial interest within the MR research community, important existing and potential applications, and major advances toward widespread clinical use. Importantly, the current status of development and remaining challenges are different for these two methods, which helps illustrate the diversity in the field of qMR.

Commented [A5]: R1.2

Commented [A6]: R1.9

Commented [A7]: R1.4

Liver Proton-Density Fat Fraction Quantification: Proton-density fat fraction (PDFF) has been developed, validated, and applied for the assessment of tissue triglyceride concentration (8). Although chemical shift encoded (CSE) fat-water imaging was introduced nearly 40 years ago (9), the development of quantitative techniques that measure PDFF has accelerated over the past two decades (10-17). Using either MRS (16) or MRI (8) acquisitions, PDFF measures the concentration of MR-visible triglyceride protons relative to all MR-visible protons (from triglycerides and water), which has multiple research and clinical applications. Recent technical developments and validation studies (see below) have led to widely available techniques. These techniques are particularly promising for the quantification of liver fat, e.g., in the assessment of non-alcoholic fatty liver disease (NAFLD).

Cardiac T_1 Mapping: Although cardiac T_1 mapping was first developed in the 90s (18), the field accelerated more recently when the promise to enable non-invasive assessment of diffuse fibrosis emerged (19). As the T_1 relaxation time depends on the mobility in the macromolecular

environment, over time cardiac T_1 mapping has proved useful in many clinical applications (20). Initially, semi-quantitative relaxation measurements in the myocardium based on Look-Locker sequences were explored (21,22). However, these methods lacked the reproducibility and reliability to serve as a quantitative tool in clinical application. With the introduction of the Modified Look-Locker Inversion Recovery (MOLLI) (23) and shortened MOLLI (shMOLLI) (24) methods, myocardial T_1 mapping became feasible on a voxel-by-voxel basis in a single breath-hold and with high visual T_1 map quality. This facilitated the widespread use and application to numerous ischemic and non-ischemic cardiomyopathies (25,26). Continuous method development and refinement have led to increasingly sensitive and reliable T_1 measurements of the heart, paving the way for routine clinical use (20).

TECHNICAL PERFORMANCE OF QMR METHODS

The development and validation of qMR methods requires a framework for describing their technical performance. The two major technical performance properties of a qMR method (Figure 2) are: i) the bias, which includes the properties of linearity, the regression slope and intercept, and the fixed bias, and ii) the precision, which is described by the repeatability and reproducibility. These metrics are described in detail below and summarized in Table 1. Previous works in qMR have used deviating terminologies, including "accuracy" or "robustness," to describe technical performance. However, in this work we use, and encourage others to use, metrics based on bias and precision, as established by the quantitative imaging metrology community (27,28). A glossary of the terminology used throughout this paper can be found in Supporting Information Table S1.

Commented [A8]: R2.4

Commented [A9]: R2.5

Bias

Bias describes the systematic tendency of qMR measurements to differ from the ground-truth value of the measurand (i.e., the underlying quantity of interest). To define bias, let X_i denote the ground-truth value of the measurand for the i-th subject, and Y_{ij} denote the j-th qMR measurement for the i-th subject. Ideally, the measurements and the ground truth possess an affine linear relationship, as follows (29,30):

$$Y_{ij} = \beta_0 + \beta_1 X_i + \epsilon_{ij} \quad (1)$$

where β_0 is the intercept, β_1 is the regression slope, and ϵ_{ij} is a random effect, which we assume is independently and identically distributed from a normal distribution with mean zero and variance σ^2 (which captures the precision).

Commented [A10]: R2.6

To measure bias, the ground-truth value can sometimes be ascertained by using a reference method. Although the ground-truth should be conceptually well defined, its estimation via a reference method is generally imperfect and requires careful design. Such a reference method may be invasive or non-invasive, and based on MR or other modalities. Importantly, a reference method should be independent from the qMR method under evaluation (e.g., should not be obtained from the same source data), as any dependence between the two measurements may lead to an underestimation of the qMR method's bias. Furthermore, to be accepted as a reference method, its measurements must be highly concordant with the ground truth, and its performance (bias and precision) must be substantially better than the performance of the method under evaluation (28). These requirements often complicate the acquisition of a reference method in vivo.

Commented [A11]: R1.5

For this reason, investigators often rely on reference objects ("phantoms") to assess bias. The design of phantoms is driven by the technique they will be testing and the specific tissues or MR properties they will mimic, as well as additional considerations such as traceability and long-term stability (31). It is important that phantoms themselves are systematically measured prior to use, which is sometimes achieved using gold standard NMR measurements (32), the best available reference method on MRI systems, or non-MR methods. With a standard phantom, such as the ISMRM/NIST system phantom (33), or a well-characterized home-built phantom, the technical performance of a qMR method can be estimated, as an approximation of in vivo technical performance. Although phantoms are highly effective in many qMR applications, there are cases where phantoms may be of limited value, as existing phantom designs do not adequately replicate the relevant signal properties, spatial distribution, or temporal dynamics found in tissue (34).

In a phantom study (or in vivo, if a suitable reference method is available), measurements are obtained at multiple values X_i (e.g., corresponding to different phantom compartments) over the range of the true value, X. Ideally, at least 10 nearly equally spaced values X_i should be chosen (29), covering the range of values of interest. This range usually includes the normal range expected in a healthy reference cohort as well as values observed under influence of the pathology or condition of interest. For each value i, the individual bias or % bias is calculated as:

Commented [A12]: R1.6, R2.10

$$b_i = (\overline{Y}_i - X_i)\,; \qquad \%b_i = \left[(\overline{Y}_i - X_i)/X_i\right] \times 100 \qquad (2)$$

where $\overline{Y_i}$ is the mean over the potentially repeated measurements on the same phantom compartment (or subject). Finally, some qMR methods may present a constant bias that is not dependent on the true value of the measurand. In these cases, the fitted line in Eq. 1 will be parallel to the identity line, with regression slope β_1 close to one, and regression intercept β_0 that provides an estimate of the bias (29).

Commented [A13]: R2.8

Over N observations, we can estimate the fixed bias:

$$\hat{b} = \sum_{i=1}^{N} \%b_i / N \qquad (3)$$

A 95% confidence interval (CI) for the fixed bias should be reported with the estimate of bias (35). Note that small values of fixed bias are often well tolerated. For example, under typical conditions, confidence intervals for a new patient's measurement constructed under the no-bias assumption provide nominal coverage as long as the fixed bias is <12% of the wSD (36).

Commented [A14]: R2.8

Because the bias sometimes depends on the true value of the measurand, the bias profile can be evaluated by plotting the estimate of bias from each value of X_i (i.e., b_i) against the true values X_i. Note that the relationship between measurements and ground truth may generally be nonlinear, particularly when considering a broad range of measurand values. However, the assumption of linearity is often an appropriate approximation and greatly simplifies the statistical analysis. To assess the property of linearity, we fit an ordinary least squares (OLS) regression of the Y_{ij}'s on X_i's. One way to test the appropriateness of a linear model is to formally test for significant non-linearities (curvature) (29). Sequential tests can be performed starting with a third-order (cubic) regression: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$. If the third order coefficient β_3 is not significantly different from zero, then the process can be repeated with a second-order (quadratic) regression: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$. If the quadratic term β_2 is not significantly different from zero, then the hypothesis of a linear model cannot be rejected, and a linear fit can be used: $Y = \beta_0 + \beta_1 X$ (29).

Commented [A15]: R2.8

Ideally, R-squared (R^2) will be greater than 0.90 (35). The regression slope, β_1, should be reported along with its 95% CI. As a general rule, a slope in the range [0.95, 1.05] is acceptable (35).

Commented [A16]: R2.9

Sometimes, the linear relationship in Equation 1 holds only for a certain range of the values of the measurand, so it is important to assess this property over the likely values of the measurand. It may be that linear relationships hold for various ranges of the true value, but β_0 and β_1 differ for each range or even vary by subject characteristics (e.g., age or body mass index).

Finally, once the bias is known, the qMR method could, in principle, be calibrated to the reference to eliminate bias. However, this is not a common approach. Indeed, the bias itself often arises from uncorrected confounding factors that may affect various acquisitions or patients differently (37,38). For this reason, bias is often not reproducible, and calibration-based correction should be approached with caution.

Precision

Precision describes the tendency of the measurement system, when used repeatedly in several "replicate" measurements on the same subject, to produce different values (27). The precision of a method has enormous practical importance. Indeed, the required number of participants for clinical studies increases with $\sigma^2$and is therefore driven by the precision of a qMR method, which determines their cost and feasibility (39). The precision is also a major factor affecting the method's ability (including sensitivity and specificity) to detect a specific condition, and determines the minimum detectable change (see Figure 3). In contrast to bias, the evaluation of precision does not require a reference method, and therefore in vivo evaluation of precision is often highly feasible.

Some studies use spatial variability in a homogeneous phantom or tissue as a heuristic to evaluate precision metrics. While this may be an acceptable approximation with certain simple imaging methods, spatial variability of system properties (e.g. B_0 and B_1^+ heterogeneities) often render this approximation inadequate, even if the region of interest appears homogeneous to the observer. In particular, this spatial variability method cannot be used to evaluate precision metrics if spatial information is used in the image reconstruction or parametric mapping (e.g., regularization in compressed sensing). Instead, precision metrics should be evaluated in studies that obtain and compare multiple replicate measurements.

Test-retest studies allow estimation of precision. When the same MR system and experimental conditions (including acquisition parameters) are used for all replicate measurements on the same subject over a short span of time, we refer to this as the *repeatability* condition. When the replicates are obtained under different conditions (e.g., different field strengths, different MRI vendors, platforms, or software versions, different individual scanners, different pulse sequences or acquisition parameters, different image analysis software, different readers, or long delay between acquisitions), we call this the *reproducibility* condition [3]. With qMR, we often characterize

Commented [A17]: R1.7, R2.34

precision by either the *within-subject standard deviation*, denoted wSD, or the *within-subject coefficient of variation*, denoted wCV. Note from Equation 1, $wSD = \sqrt{\sigma^2}$, and $wCV = \sqrt{\sigma^2}/\overline{Y}_i$. The wCV is often used when the variability in the measurements is much higher for large true values or when the measurements are log-normally distributed (thus, X_i and Y_{ij} in Equation 1 would need to be measured on a logarithmic scale).

Precision studies are often small because of cost, ethical, and technical concerns (29,30,40). For these reasons, meta-analysis is often required to pool estimates from multiple studies (41). A general rule of thumb to obtain a reliable estimate of precision is >35 subjects with two or more replicates (36).

For each subject in a test-retest study, the qMR measurement is performed at time point 1 (denoted Y_{i1}) and time point 2 (Y_{i2}). Additional time points can be included, if available. For each subject, we can calculate the mean and SD of the measurements: $\overline{Y}_i = (Y_{i1} + Y_{i2})/2$ and $SD_i = \sqrt{(Y_{i1} - Y_{i2})^2/2}$. From the N subjects, we estimate the mean wSD or wCV as:

$$\widehat{wSD} = \sqrt{\sum_{i=1}^{N} SD_i^2/N}; \qquad \widehat{wCV} = \sqrt{\sum_{i=1}^{N} \left(SD_i^2/\overline{Y}_i^2\right)/N} \qquad (4)$$

Importantly, 95% CIs for wSD and wCV should also be reported (35). Implicit in Equation 4 is the assumption that wSD (or wCV) is constant over the range of measurand values. This assumption should be assessed by calculating the estimates for several ranges of $\overline{Y}_i$ or even for various patient and/or disease characteristics, to determine a precision profile (42,43).

Two useful precision metrics are the repeatability coefficient (RC and $\%RC$), estimated as:

$$\widehat{RC} = 2.77 \times \widehat{wSD}; \qquad \widehat{\%RC} = 2.77 \times \widehat{wCV} \times 100 \qquad (5)$$

and the reproducibility coefficient (RDC and %RDC), estimated analogously. These metrics describe the smallest significant difference between two repeated measurements on a subject and assuming a normal distribution for the replicate measurements (27,30). These metrics can be used as thresholds to discern between differences due to measurement imprecision and differences due to a true change in the measurand.

In evaluating reproducibility, the experimental conditions across replicate measurements can be altered in various different ways (see above). Ultimately, the widespread dissemination of a

qMR method will require establishment of reproducibility across conditions such as different centers, MRI vendors, or patient populations. However, such multi-center, multi-vendor studies are expensive and complex, and may not be appropriate for a newly developed qMR method. A practical approach is to evaluate reproducibility in multiple studies of increasing complexity, beginning with relatively simple studies at a single center and vendor (44), while building up toward more ambitious studies (45).

Commented [A18]: R2.11

In evaluating precision, it is important to carefully design and describe the specific procedures followed. For example, repeatability may be evaluated by performing consecutive scans within the same scanning session, in order to establish the effects due to MR system adjustments and noise. However, repeatability is often determined by scanning the subject in separate sessions over a short time interval, including repositioning and re-localizing between sessions, in order to capture additional variability due to other factors such as subject positioning (29). In general, the experimental design to study precision may be different for different methods or applications. Thus, it is important to meticulously report the parameters and conditions that are kept identical and those which may have changed between replicate measurements, to enable replication and interpretation of the results.

Examples

Liver PDFF Quantification

Liver PDFF quantification methods have been validated in multiple studies including evaluation of bias in PDFF phantoms (both commercially available and home-built) (46,47), in vivo liver imaging (48), and ex vivo livers (49). In a recent meta-analysis (48), liver PDFF had high linearity and low bias with respect to the MRS-determined reference PDFF value in 23 studies, which included a total of 1,679 subjects. Test-retest repeatability studies have also been performed (48,50). Recently, high linearity and low bias of multi-center PDFF measurements (47) has been demonstrated by shipping a phantom to multiple centers in a “round-robin” study, and evaluating measurements on the same phantom across centers, vendors, platforms, field strengths, and acquisition parameters. In addition, the reproducibility of PDFF measurements in the liver, including across field strengths and MRI vendors, has also been demonstrated in multiple studies (48,51).

Cardiac T_1

Bias and precision have been the dominating criteria in analysis of cardiac T_1 mapping methods [35]. Multiple studies have shown that different T_1 mapping methods provide varying profiles of bias and precision.

Inversion recovery-based methods have been shown to exhibit good repeatability but large biases, while saturation recovery methods have been shown to reduce bias at the cost of reduced repeatability (52-54). For example, the most commonly used myocardial T_1 mapping technique, the inversion recovery-based method MOLLI, is known to be subject to multiple confounding factors and exhibits substantial bias (55). However, given its excellent repeatability and visual image quality, the sequence is highly popular among users (23). Even though some studies have shown initial evidence of multi-center or multi-vendor reproducibility with tightly controlled protocols (56), the reproducibility is generally compromised due to the measurand confounders (see next section). Thus, it is recommended to obtain center and protocol specific reference ranges in healthy subjects, before using MOLLI for quantitative diagnosis (20).

Table 1: Technical performance metrics.

Metric	Definition	Liver PDFF	Cardiac T_1 (MOLLI)
Linearity	Ability to provide measurements that are proportional to the true value as described in Equation 1	r^2 =0.96 and no evidence of significant higher-order terms in regression analysis of measurements vs true value (48).	r^2 = 0.996, magnitude of higher order terms < 0.0001(52).
Regression slope	β_1 in Equation 1	0.975 (48).	0.919 (52).
Fixed bias	Difference between average measurement and true value	<0.2% (48).	4.2% (52).
Bias (or precision) profile	A table or figure illustrating the estimates of the bias (or precision) over the range of true values and/or other relevant characteristics	See Yokoo et al., 2018 (48).	See Roujol et al., 2014 (52,57).
Repeatability	A measure of precision describing the variability in measurements on a subject over a short period of time using the same imaging system and experimental conditions (27)	Repeatability coefficient of 2.9% (48).	Repeatability coefficient of 2.0% (52) - 4.6% (58)
Reproducibility	A measure of precision describing the variability in measurements on a subject using different experimental conditions (different systems, and/or pulse sequence parameters, and/or measurements separated by a long period of time, etc) (27)	Reproducibility coefficient (across different hardware systems or reconstruction software) of 4.3% (48).	Highly variable. In tightly controlled studies, 2.1% has been reported (59), however, a meta-analysis showed >7% reproducibility in healthy subjects (60). For this reason, it is not recommended to compare MOLLI T_1 values across systems and parameters, due to system specific biases (20).

Commented [A19]: R2.34

CONFOUNDING FACTORS IN MR

In qMR, a wide variety of confounding factors may introduce bias or poor precision. Table 2 provides illustrative categories and examples of qMR confounding factors. A poll distributed among the members of the ISMRM Quantitative MR Study Group queried the frequency, relevance, and potential correction mechanisms for confounding factors used in the quantitative MR community. Supporting Information Figures S1-S6 summarize the poll results.

Commented [A20]: E.2

Table 2: Types of qMR confounding factors and illustrative examples

Hardware/system imperfections	Physiological effects/motion	Signal model imperfections	Other artifacts and noise
B_0 heterogeneities and off-resonance	Respiratory motion	Additional relaxation mechanisms (not included in model)	Partial volume
B_1 heterogeneities	Cardiovascular motion/pulsation		Slice profile imperfections
Eddy currents	Intestinal peristalsis	Spectral complexity (additional resonances, J-coupling, etc.)	Imperfect spoiling
Gradient nonlinearities	Bulk body motion		Parallel imaging artifacts
System drift	Blood flow	Exchange (multi-pool)	Noise

Commented [A21]: R2.35

Commented [A22]: R2.35

Hardware and system imperfections

The presence of magnetic field heterogeneities (both B_0 and B_1) (61), gradient nonlinearities (62), concomitant gradients (63,64), eddy currents (65), system drifts (66), timing errors (67), and other system imperfections, is unavoidable in MR applications. These effects may result in tolerable artifacts in qualitative MR as long as the relative visual contrast between tissues is preserved, but may introduce substantial bias and poor precision in qMR methods.

Commented [A23]: R2.12

Commented [A24]: R2.13

Physiological effects and motion

Physiological motion effects include respiration, cardiovascular motion and pulsation, intestinal peristalsis, and bulk patient motion, among others. These effects often result in artifacts, ghosting,

Commented [A25]: R2.14

and mis-registration in the acquired images (68), which can in turn introduce bias and poor precision in qMR. Blood and tissue motion during the acquisition can also introduce artifacts, phase offsets, and dephasing that confound the quantification.

Signal model imperfections

Practical qMR methods rely on simplified signal models. The presence of signal effects that are not included in the model introduces bias in qMR measurements. These effects may be due to additional relaxation mechanisms, incomplete approach to steady state, diffusion, spectral complexity, etc. Signal model imperfections can lead to poor reproducibility in qMR, as these effects will often manifest differently for varying experimental conditions, including different systems, field strengths, and acquisition parameters. Ideally, signal models should be based on specific biophysical assumptions about the tissues of interest. However, biophysical modeling is challenging in certain applications, and so signal "representations" are often used, which enable fitting of the acquired data but are not based on specific tissue models (69). For example, the diffusion tensor representation provides a useful approximation to the diffusion-weighted MR signal at moderate *b*-values, but is not based on specific tissue modeling assumptions (69). Such signal representations have demonstrated clinical value, but their quantitative performance needs cautious consideration. For example, bias may not be meaningful if the measurand does not represent a physical property of the tissue, and reproducibility across changes in acquisition parameters is often challenging.

Other artifacts and noise

A variety of additional imaging artifacts, including partial volume, slice profile imperfections, imperfect spoiling, parallel imaging artifacts, and noise can confound qMR methods. For example, imperfect slice profiles due to finite-duration excitation pulses lead to a distribution of flip angles across the slice and may also introduce crosstalk between slices (70,71). In addition, noise in the acquired imaging data propagates into the subsequent qMR measurements. The propagation of noise is generally dependent on the acquisition parameters and the choice of signal model; more complicated models with many free parameters often result in higher noise amplification. Further, manipulation of MR signals prior to qMR measurement can affect the noise distribution, which affects the bias and precision of qMR methods. For example, noise in complex MR data is well

modeled by a Gaussian distribution. However, qMR sometimes relies on magnitude images (e.g., in diffusion MRI, or cardiac T_1 mapping as discussed throughout this paper). This magnitude operation has several important effects, including the elimination of phase information, and the introduction of an additional bias. Indeed, regions of low signal magnitude, as commonly observed in methods such as diffusion MRI or relaxometry, deviate substantially from a Gaussian noise distribution (72). If subsequent qMR processing implicitly assumes a Gaussian noise distribution (e.g., in methods that rely on least-squares fitting, as described below), bias and poor reproducibility may result from the inaccurate noise assumptions (73,74).

For these reasons, different processing pipelines of the same data can lead to differences in the resulting qMR measurements. For example, even filtering of the image data can introduce biases in the quantification when nonlinear models are being used. Thus, using transparent open-source toolboxes or custom-built processing for qMR should be preferred over black-box tools, when reproducibility is targeted.

Finally, when using MR-based reference methods for validation of qMR, even the reference method itself may not be immune to the presence of confounding factors such as physiological effects and motion. This limitation of the reference method may complicate the evaluation of qMR bias in vivo.

Commented [A26]: R1.8

Examples

Liver PDFF Quantification

Quantification of PDFF is affected by multiple confounding factors, including:

- **T_1 recovery:** The short T_1 relaxation time of fat compared to that of water in the liver can lead to bias (overestimation) of PDFF (12) in acquisitions that include T_1 weighting.
- **T_2* relaxation:** T_2* decay across multiple echoes can appear as interference between fat and water signals, and therefore can introduce bias and poor reproducibility in PDFF quantification (see Figure 4) (13,14,75,76).
- **Spectral complexity of fat signals:** Unlike water signals, which result in a single MR resonance, fat signals arise from protons located in various positions within the triglyceride molecule. These protons, in turn, lead to a multi-peak spectrum from fat (15,77). If unaccounted for, this spectral complexity leads to bias and poor reproducibility across acquisition parameters, particularly the echo time combination.

Phase errors: Phase errors, such as those arising from eddy current effects, can introduce bias and poor reproducibility in PDFF quantification (78,79).

Over the past two decades, these and other confounding factors have been systematically identified, characterized and addressed using various acquisition and post-processing-based approaches (see the Technical Development section below).

Cardiac T_1

Quantitative cardiac imaging is particularly challenging due to the impact of cardiac and respiratory motion. Confounders, as described above, often have an imaging method-specific impact on the quantification. Additionally, subject-specific effects are often substantial. Various existing and emerging technical developments seek to mitigate these effects (55,57,80-83). The most relevant confounders for cardiac T_1 mapping include:

Heart rate: Acquisitions commonly need to be synchronized with the heartbeat to minimize cardiac motion effects. The subject-specific heart rate may influence the bias and/or precision. Introduction of alternative mapping schemes or heart rate-resilient timing has helped to alleviate this confounder (55,57,80).

k-Space acquisition: In some methods, only the effect of the magnetization preparation is modeled. In this case the disruption of the magnetization by RF pulses used for the k-space acquisition can cause bias in the quantification. Specifically, this may render the quantification susceptible to physiological factors such as T_2 relaxation (57) or magnetization transfer (84), or system-related properties such as off-resonance (85) or flip angles (86). The dependency on system-related parameters makes it paramount to use identical sequences and sequence parameters in cardiac T_1 mapping, when reproducibility across centers and MRI scanners is desired.

Partial-volume effect: Cardiac acquisitions are commonly limited in the achievable resolution due to motion constraints. This results in partial-voluming, where voxels are partially filled with different tissue types at tissue interfaces. As a result, the area that can reliably be evaluated is further reduced, rendering the quantification dependent on accurate delineation of the region of interest (87,88).

Commented [A27]: R2.16

Commented [A28]: R2.17, R2.18

Commented [A29]: R2.19

TECHNICAL DEVELOPMENT AND VALIDATION

The development of qMR methods is typically an iterative process including design of acquisition, modeling and signal fitting methods. This technical development can be framed as an optimization of bias and precision metrics (see Figure 5), subject to specific constraints such as scan time, or hardware performance.

Commented [A30]: R2.20

Acquisition

Once the basic physical mechanism to be probed has been selected and potential confounders have been identified, aspects of protocol design and optimization can be considered. A common goal is to select pulse sequences and parameters such that the measurand can be determined with low bias and high precision, subject to a set of timing, hardware, and other constraints. Acquisition design will often begin by selecting a pulse sequence where the measurand of interest can be directly probed, while minimizing the effect of confounding factors. Next, an acquisition that includes multiple scans with different parameters can be designed to enable estimation of the measurand. In applications where thermal noise is the dominant source of noise (as opposed to, e.g., physiological noise), acquisitions with higher imaging SNR may substantially improve bias (12) or precision (89). The choice of acquisition parameters may be driven by heuristics, and also refined using quantitative tools such as sensitivity analysis (90,91), or noise propagation analysis (e.g., Cramer-Rao lower bounds, CRLB) (11,92-94).

Commented [A31]: R2.22

Liver PDFF Quantification

The choice of pulse sequence for quantification of PDFF is driven by the desire to obtain chemical shift-encoded data with proton-density contrast (e.g., avoiding confounding effects due to T_1 and T_2 relaxation), and with rapid scan times (e.g., to enable whole-liver coverage in a single breath-hold while avoiding motion artifacts). For these reasons, the pulse sequence of choice for MRI-based liver PDFF quantification is a multi-echo spoiled gradient echo (SGRE) sequence, either using 2D multi-slice or 3D imaging (8,17). In addition, small flip angles are used in order to avoid T_1 bias (12). Other confounding factors are typically addressed by postprocessing/modeling (see below). Optimal acquisition parameters, such as echo times, have been determined using CRLB analysis (11).

Cardiac T_1

A multitude of pulse sequences for cardiac T_1 mapping have been proposed and novel method development remains an active area of research. Generally, these acquisitions can be decomposed into three integral parts: 1) contrast sensitization; 2) k-space acquisition; 3) motion compensation.

Typically, preparation pulses are used to sensitize the imaging signal to the T_1 relaxation time of the tissue. Inversion pulses are most widely used in T_1 mapping, including the commonly used MOLLI sequence and its variants (23,24,55). Saturation recovery has also been proposed with the potential to minimize bias caused by various confounding factors as described above (57,80). However, due to a decreased dynamic range, saturation recovery preparation typically results in lower T_1 mapping precision as compared with inversion preparation.

Cardiac T_1 mapping is typically performed using multiple electrocardiogram (ECG)-triggered snapshot images, with each image obtained during a single diastolic quiescence, i.e., all *k*-space lines necessary for image reconstruction of one snapshot image are acquired in one heartbeat. To achieve optimal SNR as well as minimal disruption of the longitudinal magnetization recovery curve, balanced Steady State Free Precession (bSSFP) readouts are the method of choice. Spoiled gradient echo readouts have also been explored to minimize sensitivity to off-resonance and field heterogeneities, albeit at the cost of reduced precision (95). More recently, continuous imaging throughout the heartbeat have been proposed to allow cardiac phase-resolved T_1 mapping (81,82,96).

Commented [A32]: R2.17

ECG triggering is almost universally used as the means for cardiac motion compensation in T_1 mapping, with few notable exceptions (96,97). Various schemes have been explored for respiratory motion compensation. Clinically available T_1 mapping methods usually acquire a single-slice T_1 map in a single breath-hold (26). However, free-breathing methods have also been explored with diaphragmatic navigator gating, tracking or self-gating (97-99). Importantly, free-breathing sequences allow for the acquisition of multiple slices or 3D volumes and can be used to enable T_1 mapping with increased spatial resolution (83,100,101).

Model selection

Many qMR methods rely on parametric mapping using a signal model that relates the acquired data to the underlying measurand. Selection of a signal model is typically an iterative process and seeks to balance bias and precision. Often the process begins with identifying the relevant degrees

of freedom in the underlying tissue (69), such that these tissue properties can be related to, and estimated from, acquired MR signals. For example, this step may involve the identification of the major pools of nuclei with shared properties that will reasonably contribute to the signal. These pools can describe physical compartments, such as "intracellular" compartments, or local molecular environments such as lipid protons. The Bloch equations, describing the response to RF energy and relaxation properties, are defined for each pool. Next, a model may consider whether nuclei can travel between pools by chemical exchange or diffusion, or interact magnetically with other pools due to proximity and thus define the exchange kinetics. Models commonly describe the signal within a voxel independently of its spatial neighborhood, but one may also need to consider the influence of the neighboring voxels (e.g., in quantitative susceptibility mapping (102), or electrical properties tomography (103,104)).

The next step is often to evaluate the signal model under modifications of the acquisition pulse sequence. It is often helpful to develop a working model for simple excitation-readout with Cartesian acquisition and long TR before considering advanced *k*-space trajectories or pulse trains. The requirements for each measurand are different, but major considerations in the presence of increasingly advanced pulse sequences may include relaxation effects, B_0 and B_1 heterogeneities, etc. It may also be necessary to consider the need for steady-state or non-steady state modeling. Finally, any signal manipulations that occur before analysis, such as magnitude operation or spatial filtering, need to be included.

In subsequent iterations, confounding factors are often identified and addressed through acquisition- and/or modeling-based refinements. Importantly, qMR methods necessarily use simplified models of the actual underlying physics. For this reason, it is always possible to "enhance" the models by including additional unknown parameters. However, these signal model enhancements generally lead to increased challenges in the parameter estimation (particularly noise amplification, sensitivity to artifacts, and computation time). Practical signal models, therefore, seek a balance between accurately capturing the underlying physics and enabling stable quantification within acceptable computation times. Once a satisfactory model is achieved, this model often needs to be re-evaluated upon subsequent refinements of the qMR method, including accelerated acquisitions.

Liver PDFF Quantification

A widely used signal model for PDFF quantification in the liver includes (8,15,105):

$$s(TE_n) = \left(\rho_w + \rho_F \sum_{m=1}^{M} \alpha_m e^{i2\pi f_{F,m} TE_n}\right) e^{i(\phi_0 + 2\pi f_B TE_n)} e^{-TE_n/T_2^*} \quad (6)$$

where ρ_w and ρ_F are the proton density-weighted signal amplitudes of water and fat, respectively; fat signals are modeled as a pre-calibrated spectrum including M peaks with known relative amplitudes α_m and frequency offsets $f_{F,m}$ (77); initial phase ϕ_0; B_0 related off-resonance frequency f_B; and transverse relaxation time T_2*. Upon data fitting (see below), this signal model allows estimation of ρ_w and ρ_F, which lead to the calculation of PDFF as:

$$PDFF = \frac{\rho_F}{\rho_w + \rho_F} \quad (7)$$

Importantly, the widely used signal model in Equation 6 constitutes a balance between bias and precision (noise performance) (106). For example, this model addresses the spectral complexity of the fat signal by using a multi-peak signal model and also accounts for T_2* decay. If unaccounted for, both of these effects have been shown to lead to substantial bias and poor reproducibility in PDFF quantification (17). However, the model in Equation 6 typically relies on a pre-calibrated multi-peak fat spectrum, where the relative frequencies and amplitudes of the fat peaks are assumed known a priori (107), and also assumes a common T_2* decay time for water and all fat peaks (106). Each of these approximations help maintain acceptable noise performance and precision for PDFF quantification by limiting the number of unknown parameters, even though they may introduce a small bias in the estimation of liver PDFF when the model assumptions do not hold exactly.

Cardiac T_1

T_1 recovery is thoroughly studied and can be accurately described by the well-known phenomenological Bloch relaxation equations. However, the signal model needs to be adapted to the specific imaging sequence, as summarized next.

In the widely used Modified Look-Locker Inversion Recovery (MOLLI), an adaptation of the standard inversion recovery model is commonly employed to describe the signal $S(t)$ at different inversion times t:

$$S(t) = A - Be^{-t/T_1^*}. \quad (8)$$

Here A and B describe fit parameters and T_1^* is the apparent relaxation time. The T_1 estimate is then extracted as $T_1 = \left(\frac{B}{A} - 1\right)T_1^*$. This adaptation is inspired by Deichmann et al. (108) and aims to reduce the effect of the RF pulses used for the k-space acquisition on the quantification. However, the acquisition commonly deviates from the assumptions underlying this correction, inducing residual susceptibility to various effects related to the RF pulses during the k-space acquisition. Numerical models have also been proposed to approximate the magnetization evolution during the k-space acquisition, using for example Bloch equation simulations or additional parameters (109,110). While these methods commonly achieve lower bias, their applicability might be limited, and precision may be compromised.

Commented [A33]: R2.17

Commented [A34]: R2.17

Commented [A35]: R2.25

A second class of myocardial T_1 mapping methods is based on saturation recovery, which commonly relies on the following three-parameter model:

$$S(t) = A\left(1 - e^{-t/T_1}\right) + B \qquad (9)$$

This saturation recovery-based approach has been shown to compensate for the effects of the RF pulses used for the k-space acquisition (57), which in turn enables cardiac T_1 mapping with reduced bias. It has also been suggested to omit the B parameter to obtain a two-parameter model, in order to improve precision at the cost of bias in saturation recovery T_1 mapping (55).

Commented [A36]: R2.17

Model Fitting

Fitting a signal model to acquired data can typically be described in terms of a formulation and an algorithm, as described next.

The formulation is often an optimization problem, which describes in what sense the model should fit the acquired data. For example, least-squares fitting is often used for various linear or nonlinear models in quantitative MR. For nonlinear models based on exponential functions, a logarithm of the data is sometimes calculated to linearize the problem. This linearization simplifies the optimization, although it affects the noise propagation and may require additional manipulations to avoid excessive noise influence from low-SNR data points (111). Further, some formulations rely on the acquired complex data, whereas others use magnitude data (79). Finally, the formulation may be constrained (where the set of allowable parameters is restricted based on physical or noise propagation considerations) or unconstrained. In addition to least-squares fitting,

other formulations can be used, including those required for maximum-likelihood estimation in the presence of non-Gaussian noise.

Once a formulation is selected, an algorithm needs to be selected to solve the corresponding optimization problem. Depending on the formulation, various closed-form or iterative algorithms are typically available. Variations of Newton's method, including Levenberg-Marquardt and Gauss-Newton algorithms, constitute common choices for iterative optimization (112). An ideal algorithm would be efficient (i.e., fast and requiring low resources) and would lead to the global solution of the optimization problem described in the formulation.

Liver PDFF quantification

Model fitting for PDFF mapping is typically performed using nonlinear least-squares fitting of the signal model (Eq. 6) to the acquired multi-echo data, followed by calculation of PDFF at each pixel (Eq. 7). Multi-echo data are often corrupted by phase errors that are inconsistent across echoes (78,79). For this reason, some or all of the phase information is often discarded to avoid PDFF bias, and algorithms often rely partly on fitting the signal magnitude, instead of the original complex-valued signals. Magnitude fitting leads to reduced bias by avoiding phase related PDFF errors at the cost of reduced noise performance and precision (by discarding half of the acquired information, i.e., the phase).

Cardiac T_1

Basic model fitting in cardiac T_1 mapping is also most commonly performed using magnitude-based nonlinear least-squares fitting. When unsigned magnitude images are used in an inversion-recovery model, the signal polarity information is lost. To resolve this issue the images are commonly ordered by the inversion time and the polarity can be restored heuristically, by successively flipping the sign in the ordered sequence and accepting the solution with the lowest fit residual (20).

However, this process might introduce additional noise variability. It has been proposed to incorporate phase information to perform hybrid fitting on a signed magnitude. Here the background phase is extracted from a fully relaxed image, and the phase difference to other T_1 weighted images can be used to restore the signal polarity (113).

CLINICAL QUALIFICATION

For qMR methods, in addition to technical validation, which measures the bias and precision of the quantitative measurements, it is essential to perform clinical qualification (Figure 6) (114). Clinical qualification seeks to establish the relationship between the qMR measurement and biological processes or clinical endpoints, as needed to determine the clinical utility of the method, e.g., whether it enables screening, diagnosis, staging, prognosis, or treatment monitoring for a particular condition and target population (115-119). For example, rather than focusing on technical performance metrics of bias and precision, clinical qualification may focus on metrics such as sensitivity, specificity, negative / positive predictive value, prediction accuracy, or odds ratio (114,120). Upon successful clinical evaluation for a specific application, qMR methods may lead to qualified quantitative biomarkers (see Glossary in Supporting Information Table S1) (6).

This marks an important distinction: while a qMR measurand usually relates to a physical property, this measurand, when being used as a clinical biomarker, indicates pathophysiological alterations or other changes in the physiological state. Often numerous biological and physiological processes affect the underlying physical property. Thus, a single qMR measurand can be qualified as a biomarker for multiple disease entities. In this case, while being sensitive to multiple diseases, the measurand may not be specific to any one physiological alteration.

There are strong connections between technical validation and clinical qualification. For example, a qMR method with poor precision (e.g., poor test-retest repeatability) will likely also have poor sensitivity and specificity for detection of a specific condition. However, these are also important distinctions between both types of evaluation. For example, it is possible to develop a qMR method with excellent technical performance (low bias and high precision) for quantifying a measurand; however, this method may have poor clinical performance for a specific application, e.g., due to underlying biological variability that complicates the relationship between the measurand and the clinical endpoint of interest, such as survival, disease-free survival, or various surrogate endpoints (121-123). Further, a biased method may reduce the desired effect size, as the bias itself may be different for various patient populations. Alternatively, it is possible that a biased (confounded) qMR method provides larger effect sizes for a specific disease entity than an unbiased measurement, e.g., if the confounders themselves are sensitive to the physiological alteration (see Figure 7). However, it is important to note that this enhancement usually comes at the cost of strongly reduced reproducibility as the variability in the bias is difficult to control.

Commented [A37]: P2.26

Commented [A38]: E.2

Commented [A39]: R2.28

Commented [A40]: R2.29

For these reasons, it is essential to conduct both technical validation and clinical qualification of qMR methods. This need further highlights the importance of multi-disciplinary collaboration between technical imaging researchers, translation-focused radiologists, and other clinicians.

Examples

Liver PDFF Quantification

Liver PDFF has been shown to be correlated with histologic steatosis grade. For example, PDFF can classify histologic steatosis (grade 0 vs. 1-3) with sensitivity 0.93 and specificity 0.94 (124). Also, MRI-based liver PDFF quantification is emerging as a useful biomarker to assess longitudinal changes in liver fat within clinical trials (125). Further, a reduction of MRI-PDFF by 30% is associated (odds ratio 6.98) with histologic improvement in NAFLD Activity Score (126).

Commented [A41]: R2.30

Liver PDFF values may be predictive of pediatric metabolic syndrome (127). In addition, liver steatosis is associated with cardiovascular diseases (128,129); for instance, liver fat is an independent risk factor (odds ratio 2.1) for high-risk plaque (128) and other cardiovascular risk factors (129). Importantly, now that the required MRI technical development is mature and PDFF mapping methods are widely available, determination of the association between liver PDFF and various clinical outcomes constitutes an active area of research.

Cardiac T_1

Native myocardial T_1 times, in the absence of a contrast agent, have been evaluated against histologically determined fibrosis from myocardial biopsies in vivo and total collagen volume in animal studies (130) and heart transplant patients (131). Variable degrees of correlation ranging from moderate to high have been reported depending on the disease model (131), indicating that T_1 times are not reflective of fibrosis alone but of a number of factors. Extracellular Volume (ECV) calculated from native T_1, post-contrast T_1, and hematocrit, generally showed better correlation to the amount of fibrosis but variability among disease models and studies remains. Accordingly, the clinical context needs to be considered in the interpretation of both native T_1 and ECV, and alteration in either measurand cannot be directly linked to a single specific physiological process (20,132-134).

Nonetheless, cardiac T_1 mapping-related markers have demonstrated high clinical diagnostic and prognostic value in an unexpectedly wide range of disease entities (25,135). For example, in cardiac amyloidosis ECV showed excellent sensitivity and specificity (0.93 and 0.87, respectively) and an odds-ratio of 84.6 (136). In patients with an acute infarct quantitative assessment of normal appearing myocardium in patients with an acute infarct using native T_1 or ECV has proven to be a better predictor for all-cause mortality or major cardiac events than any other cardiac MRI marker (131), and accurate differentiator between reversible and irreversible myocardial damage (96.7% prediction accuracy).

Commented [A42]: R2.30

Interestingly, MOLLI T_1 mapping, which is most common in clinical use, is known to exhibit a large bias. However, it has been suggested that certain confounders to MOLLI T_1 measurements may enhance clinical sensitivity (137). As illustrated in Figure 7, this somewhat counter-intuitive phenomenon arises because some confounders (e.g., magnetization transfer (84,137,138)) are sensitive to pathological alterations, leading to inflated effect size in certain disease entities compared to unbiased measurements. However, as a result of these confounders, the measurand in MOLLI is highly dependent on the sequence parameters (e.g., TR, flip-angle, slice-profile), the scanner specifications and tissue properties that are not related to the tissue T_1 time. Hence, this inflated effect size is obtained at the cost of reduced reproducibility.

Commented [A43]: R2.31

DISSEMINATION IN RESEARCH AND IN THE CLINIC

Although qMR methods have shown great potential to guide clinical decision-making and patient management for improved patient care and outcomes, very few qMR methods are used in routine clinical practice. Many promising qMR methods are only described in the scientific literature, without translation in clinical research studies or clinical use.

Commented [A44]: R2.32

As recently stated in the *Imaging Biomarker Roadmap for Cancer Studies* (139), all imaging biomarkers, including quantitative MRI biomarkers, need to cross two "translational gaps" before they are ready to guide clinical decision making. These gaps are crossed through increasing technical validation and clinical qualification, as well as assessment of cost effectiveness and other considerations. Once technical and clinical performance evaluation demonstrate the reliability of a qMR biomarker to test medical research hypotheses, this biomarker can cross the first gap and become a useful "medical research tool". At this stage, substantial additional validation and

qualification are still needed in order to achieve clinical impact as a screening, diagnostic, or predictive biomarker. This may include validation of multi-center reproducibility (140), and large prospective clinical trials to demonstrate improved clinical diagnosis or outcomes. In combination with cost effectiveness and other considerations, these activities enable a biomarker to cross the second translational gap and become a "clinical decision-making tool" that influences patient care. Importantly, application of qMR methods in research or in the clinic requires rigorous quality assurance and quality control procedures (141-143).

During the three processes of i) technical development and validation, ii) clinical qualification, and iii) dissemination, additional substantial challenges (including regulatory issues and market-related factors) often arise before clinical dissemination is achieved (Figure 6). For example, the ability of healthcare providers to obtain reimbursement or take charge of the costs associated with quantitative imaging biomarkers may drive the clinical use of these tools. Oftentimes, a lack of CE/FDA (or equivalent) labelling limits the ability to apply a biomarker in clinical practice.

Examples

Liver PDFF Quantification

CSE-based liver PDFF quantification has emerged as a major clinical and research tool to determine liver fat content. Importantly, this qMR method is commercially available on systems from various MRI vendors, including regulatory approvals such as FDA clearance and CE mark. A liver PDFF quantification profile is currently being developed by the Radiological Society of North America's Quantitative Imaging Biomarkers Alliance (RSNA QIBA) (144). Now that liver PDFF quantification methods have shown excellent technical performance (low bias, high precision), clinically relevant results are emerging, including population studies measuring prevalence in various populations (145) and clinical studies showing the prognostic value of PDFF (146).

Commented [A45]: R2.33

Cardiac T_1

In recent years, cardiac T_1 mapping has become widely available on most clinical MRI systems. Some vendors have released dedicated product packages comprising one or more T_1 mapping methods, while others have provided prototype methods. Several cardiac T_1 mapping methods have regulatory approval such as FDA clearance and CE mark.

Commented [A46]: R2.33

T_1 mapping is widely used in cardiac MRI in academic centers and beyond. It has been successfully applied to an unexpectedly large spectrum of ischemic and non-ischemic cardiomyopathies (20,131) and is established as part of routine scanning in numerous clinical cardiac MR protocols. The effect of most heart diseases on myocardial T_1 has been investigated, mostly in single center studies. Select pathologies have been studied in large cohorts or multi-center studies, including studies on amyloidosis (147) and Anderson-Fabry disease (148). Additionally, cardiac T_1 mapping has been adopted in multiple national cohorts, including the UK biobank protocol and the German national cohort (149,150). These studies are some of the largest ongoing MRI projects to date. Following the clinical success demonstrated in the literature, cardiac T_1 mapping was adopted in disease specific clinical guidelines (151). Further increases in clinical integration and use in a growing number of cardiac MRI protocols are likely.

RELATED INITIATIVES, CHALLENGES AND OPPORTUNITIES

As described above, substantial efforts are needed for the development, validation and dissemination of quantitative MR techniques. These efforts require collaboration between technical researchers, translational researchers and clinicians, industry, and initiatives and institutions dedicated to the regulation and guidance of quantitative imaging measurements. Such initiatives and institutions include authorities for standardization of measurements such as Italy's Istituto Nazionale di Ricerca Metrologica (INRIM), the Korea Research Institute of Science and Standards (KRISS), the U.S.' National Institute of Standards and Technology (NIST), the UK's National Physical Laboratory (NPL), Germany's Physikalisch-Technische Bundesanstalt (PTB), and for the advancement of the development and use of imaging biomarkers, as performed by QIBA, the US National Cancer Institute through the Quantitative Imaging Network (QIN), the European Imaging Biomarkers Alliance (EIBALL), Japan-QIBA, the European Society of Radiology (ESR), and the ISMRM. Specifically, the major goal of the ISMRM Quantitative MR Study Group is to promulgate documentary and measurement standards for qMR methods in collaboration with national metrology institutes, academic and clinical MR sites, and through collaboration with existing study groups. Further, ongoing qMR improvements occur in the context of broad efforts to evaluate and optimize the value of MRI in medicine (152). Ultimately,

efforts to develop qMR methods should have broad value in medicine across countries and populations, beyond specialized research centers.

Commented [A47]: R1.12

In addition, the development and validation of qMR methods is closely connected to the improvement of the reproducibility of MR research itself. There are multiple existing and emerging initiatives in this area, including the ISMRM Reproducible Research Study Group, and reproducibility has recently been emphasized by major journals such as *Magnetic Resonance in Medicine,* or the *Journal of Magnetic Resonance Imaging*. A related set of benchmarks for validation of quantitative imaging tools has been described by the Quantitative Imaging Network of the US' National Cancer Institute (153). Importantly, multiple consensus efforts and community challenges have emerged in recent years for specific qMR methods or applications, as well as for general optimization of qMR (6,20,37,142,154-163). Data standards are also essential for reproducibility and interoperability, making it easier to create transparent qMR workflows. Two standards that are relevant for qMR are the ISMRM-Raw Data format (https://ismrmrd.github.io/apidocs/1.5.0/), and the Brain Imaging Data Structure (BIDS) extension proposal for quantitative MRI (https://github.com/bids-standard/bids-specification/pull/508).

Further, various software packages developed, maintained and used by the community enable improved reproducibility by standardizing data processing pipelines. Supporting Information Figure S6 gives an overview of the user base of publicly available software with applications in qMR. Examples include various toolboxes hosted on the Matlab Central File exchange (https://www.mathworks.com/matlabcentral/fileexchange/), FSLTools (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslTools), OsiriX or Horos plugins, ImageJ (https://imagej.nih.gov/ij/), Bay Area Reconstruction Toolbox (BART, https://mrirecon.github.io/bart/), qMRlab (https://qmrlab.org) (164), Gadgetron (http://gadgetron.github.io/), Quantitative Imaging Tools (https://github.com/spinicist/QUIT), Michigan Image Reconstruction Toolbox (MIRT, http://github.com/JeffFessler/MIRT.jl), hMRI (https://hmri-group.github.io/hMRI-toolbox/), , QMRI tools (https://community.wolfram.com/groups/-/m/t/1661539), LCModel (http://s-provencher.com/lcmodel.shtml), Total Mapping Toolbox (TOMATO) (https://mrkonrad.github.io/TOMATO/html), QMRI Tools (https://mfroeling.github.io/QMRITools/), and others (e.g., vendor proprietary software and in-house or personal code). RSNA QIBA and NIBIB have also sponsored the development of digital

reference objects (DROs), which enable the testing of analysis tools to assess their bias and precision when working with quantitative data obtained with different acquisition parameters and varying levels of SNR. Example DROs for DCE-MRI and DWI are available from QIBA (https://qidw.rsna.org).

The transformation of MR into a truly quantitative diagnostic modality has enormous potential to impact research and clinical care. However, the development, validation, and dissemination of quantitative MR methods is faced with multiple challenges, particularly the complexity and cost of the required validation studies, as highlighted by the above networks and initiatives. These challenges reinforce the need for collaboration between technical MR researchers, academic radiologists, and other clinicians, as well as industry, such as Original Equipment Manufacturers (OEMs) - vendors of MR systems and other MR equipment and software, pharmaceutical companies, contract research organizations, and others.

Finally, substantial recent efforts from the qMR community have focused on rapid multi-parametric mapping and machine learning (ML). Emerging multi-parametric mapping methods such as MR fingerprinting (165) and multi-tasking (81) enable quantitative mapping of several parameters with short scan times. These methods are highly promising for a variety of applications, and require careful development and validation to address a large number of potential confounding factors. ML methods, including radiomics and deep learning, have recently gained enormous interest in the field. Indeed, ML may contribute to different stages of the qMR pipeline, including image prescription, acquisition, reconstruction, post-processing, measurement, and analysis. Despite the potential impact of these methods, rigorous development and validation of ML-enabled qMR is needed. This development and validation pose new challenges and opportunities for ML-enabled qMR, including how to quantify and address confounding factors to achieve low bias and high reproducibility across patients, sites, and vendors, in much the same way as the more ‘traditional’ qMR methods highlighted in the present manuscript.

SUMMARY AND CONCLUSION

On behalf of the International Society for Magnetic Resonance in Medicine (ISMRM) Quantitative MR Study Group, this manuscript describes a framework for the development and validation of quantitative MR methods. With a focus on technical performance metrics (bias and precision), this

framework highlights the challenges as well as the research opportunities associated with quantitative MR methods. Overall, rigorous development and validation are critical components of the transformation of MR into a truly quantitative diagnostic modality. A summary of concluding recommendations to achieve this aim is provided in Table 3. Upon successful implementation of qMR methods, as well as clinical qualification of qMR-based biomarkers, qMR has the potential to substantially advance imaging in clinical applications and clinical research, and build a cornerstone of precision radiology.

Table 3: Recommendation for qMR development, validation and application.

Definitions	The measurand of interest needs to be clearly defined. How does the targeted measurand relate to other physical properties? For example, if a coefficient is determined, what is the reference quantity (e.g., MR-visible protons)?
Choice of pulse sequence	Select pulse sequences and parameters such that the measurand can be determined with low bias and high precision, subject to a set of timing, hardware, and other constraints. Acquisition design will often begin by selecting a pulse sequence where the measurand of interest can be directly probed, while minimizing the effect of confounding factors.
Choice of models	Proper biophysical modeling is difficult, but may avoid the pitfalls of various signal representations. Indeed, various models can often fit the data, but models that are not grounded on specific tissue assumptions are often more difficult to validate, and are also likely to suffer from poor reproducibility.
Rigorous validation	It is critical to perform systematic validation of the technical performance of emerging qMR methods. Importantly, even though early-stage validation is often focused on bias, evaluation of precision (repeatability and reproducibility) is essential to enable further clinical qualification and dissemination.
Structured evaluation	Well-structured reporting of the validation is an essential component of establishing a qMR method. The standard metrics being evaluated should be described clearly as discussed in the section "Technical Performance of qMR Methods" above. Future work from the community may establish a standardized structure for the Methods and Results sections of qMR manuscripts.

Commented [A48]: R2.36

Real-world validation	Even at the stage of technical validation, it is important to evaluate the performance of qMR methods under conditions that are relevant to the real-world clinical environment. For example, performance may depend on the hardware available at different sites (e.g., academic vs non-academic). In addition, technical validation in a relevant patient population helps pave the way for subsequent clinical qualification and application.
Focus on reproducibility	Optimization and characterization of reproducibility across acquisition protocols, field strength, vendor, platform, etc, is critical in qMR. Indeed, in qualitative MR methods development, one is often interested in finding the optimal set of acquisition and processing parameters to maximize imaging performance (e.g., resolution, SNR). Although this optimal set of parameters is also relevant in qMR, the development of qMR methods that are reproducible across variations in the acquisition parameters is arguably even more important than the identification of the optimal parameters. This way, qMR methods are best suited for widespread dissemination across sites that may not be able to implement exactly optimized acquisitions.
Reproducibility vs. standardization	Certain qMR methods are highly reproducible across variations in acquisition parameters (within a certain range). For example, this is the case for PDFF measurement in the liver: when correcting for all relevant confounders, PDFF measurement is highly reproducible across field strength, echo time combinations, spatial resolution, and various other acquisition parameters. However, other qMR methods have poorer reproducibility, and their widespread dissemination would benefit highly from standardization of acquisitions (as well as processing) across sites and systems, as well as harmonization (see below).
Harmonization	Quantitative MR can benefit from harmonized acquisitions and tools. For example, standardized reference objects and tool validation methods, such as the use of DROs, provide common ground for comparison of imaging protocols across sites, vendors, and software analysis packages.
Realistic time-horizon	Development, validation, qualification, and dissemination of a qMR method is a slow, iterative process that may take more than a decade.
Consider the end goal	In qMR, the end goal is often to enable improved diagnosis, staging, and/or treatment monitoring of disease and generate

	increased value in the clinical work up. This goal is generally relevant even for technically focused researchers.
Mind the translational gap	Establishing new quantitative imaging biomarkers in actual clinical use is a lengthy process and requires many steps that may not be amenable to funding by traditional science-oriented grants or tenders. This may require creativity about the path to clinical integration and real-world use.
Clinical qualification is key	Clinical qualification is critical to achieve translation of a quantitative MR method to the clinic. This may be the most time-consuming step in the entire pipeline of qMR method development and evaluation.
Collaboration	Working together with stakeholders (technical, clinical, industrial) and across imaging modalities or scientific disciplines is critical. For example, accurate biophysical modelling will benefit from collaboration between clinical and preclinical MR scientists, but also between MR researchers and scientists studying tissues at smaller scales (e.g., cell cultures) or using different imaging technology (e.g., X-ray phase contrast imaging for tissue structure, near infrared spectroscopy for blood oxygenation properties, or microscopy). Collaboration with clinicians is of enormous value in qMR technique development, and helps create a virtuous loop of refinement of existing methods and conception of new methods that address existing clinical needs. Further, early-stage discussion and cooperation with industry is especially relevant, since CE/FDA-labelling is mandatory for clinical translation. A technique without labelling will not be widely adopted in clinical practice due to ethical concerns and regulatory issues.

Commented [A49]: R1.10

ACKNOWLEDGMENTS

The authors would like to thank the ISMRM Quantitative MR Study Group Consortium, including those who responded to the online poll of the ISMRM Quantitative MR Study Group, as well as those who provided feedback during the public presentation on June 23, 2021, for their thoughtful feedback, which has led to an improved manuscript. We also acknowledge the efforts of the endorsers, as listed in the Supporting Information Table S2. We also thank the ISMRM Publications Committee for reviewing the work and coordinating the ISMRM Board of Trustees approval process.

CONFLICT OF INTEREST

Nancy Obuchowski is a paid consultant for RSNA's Quantitative Imaging Biomarker Alliance. Bettina Baessler is co-founder of Lernrad GmbH, Germany. Xavier Golay is a co-founder, CEO and shareholder of Gold Standard Phantoms. Diego Hernando is co-founder of Calimetrix, LLC.

REFERENCES

1. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69(3):89-95.
2. Abramson RG, Burton KR, Yu JP, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Lenchik L, Subramaniam RM. Methods and challenges in quantitative imaging biomarker development. Acad Radiol 2015;22(1):25-32.
3. Cercignani M, Dowell N, Tofts P, editors. Quantitative MRI of the brain: principles of Physical measurement. 2nd Edition ed. Boca Raton, FL, USA: CRC Press; 2018.
4. Cui Y, Zhang XP, Sun YS, Tang L, Shen L. Apparent diffusion coefficient: potential imaging biomarker for prediction and early detection of response to chemotherapy in hepatic metastases. Radiology 2008;248(3):894-900.
5. Abramson RG, Arlinghaus LR, Dula AN, Quarles CC, Stokes AM, Weis JA, Whisenant JG, Chekmenev EY, Zhukov I, Williams JM, Yankeelov TE. MR Imaging Biomarkers in Oncology Clinical Trials. Magn Reson Imaging Clin N Am 2016;24(1):11-29.
6. deSouza NM, Achten E, Alberich-Bayarri A, Bamberg F, Boellaard R, Clement O, Fournier L, Gallagher F, Golay X, Heussel CP, Jackson EF, Manniesing R, Mayerhofer ME, Neri E, O'Connor J, Oguz KK, Persson A, Smits M, van Beek EJR, Zech CJ, European Society of R. 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):87.
7. Modell B, Khan M, Darlison M, Westwood MA, Ingram D, Pennell DJ. Improved survival of thalassaemia major in the UK and relation to T2* cardiovascular magnetic resonance. J Cardiovasc Magn Reson 2008;10(1):42.
8. Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. Magn Reson Imaging Clin N Am 2010;18(3):337-357, ix.
9. Dixon WT. Simple proton spectroscopic imaging. Radiology 1984;153(1):189-194.
10. Reeder SB, Wen Z, Yu H, Pineda AR, Gold GE, Markl M, Pelc NJ. Multicoil Dixon chemical species separation with an iterative least-squares estimation method. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2004;51(1):35-45.
11. Pineda AR, Reeder SB, Wen Z, Pelc NJ. Cramer-Rao bounds for three-point decomposition of water and fat. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2005;54(3):625-635.
12. Liu CY, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: correction of bias from T(1) and noise. Magn Reson Med 2007;58(2):354-364.
13. Yu H, McKenzie CA, Shimakawa A, Vu AT, Brau AC, Beatty PJ, Pineda AR, Brittain JH, Reeder SB. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. Journal of magnetic resonance imaging : JMRI 2007;26(4):1153-1161.
14. O'Regan DP, Callaghan MF, Wylezinska-Arridge M, Fitzpatrick J, Naoumova RP, Hajnal JV, Schmitz SA. Liver fat content and T2*: simultaneous measurement by using breath-hold multiecho MR imaging at 3.0 T--feasibility. Radiology 2008;247(2):550-557.

15. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2008;60(5):1122-1134.
16. Hamilton G, Middleton MS, Bydder M, Yokoo T, Schwimmer JB, Kono Y, Patton HM, Lavine JE, Sirlin CB. Effect of PRESS and STEAM sequences on magnetic resonance spectroscopic liver fat quantification. J Magn Reson Imaging 2009;30(1):145-152.
17. Meisamy S, Hines CD, Hamilton G, Sirlin CB, McKenzie CA, Yu H, Brittain JH, Reeder SB. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology 2011;258(3):767-775.
18. Schwarzbauer C, Syha J, Haase A. Quantification of regional blood volumes by rapid T1 mapping. Magn Reson Med 1993;29(5):709-712.
19. Everett RJ, Stirrat CG, Semple SI, Newby DE, Dweck MR, Mirsadraee S. Assessment of myocardial fibrosis with T1 mapping MRI. Clin Radiol 2016;71(8):768-778.
20. Messroghli DR, Moon JC, Ferreira VM, Grosse-Wortmann L, He T, Kellman P, Mascherbauer J, Nezafat R, Salerno M, Schelbert EB, Taylor AJ, Thompson R, Ugander M, van Heeswijk RB, Friedrich MG. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). J Cardiovasc Magn Reson 2017;19(1):75.
21. Mewton N, Liu CY, Croisille P, Bluemke D, Lima JA. Assessment of myocardial fibrosis with cardiovascular magnetic resonance. J Am Coll Cardiol 2011;57(8):891-903.
22. Amano Y, Takeda M, Tachi M, Kitamura M, Kumita S. Myocardial fibrosis evaluated by Look-Locker and late gadolinium enhancement magnetic resonance imaging in apical hypertrophic cardiomyopathy: association with ventricular tachyarrhythmia and risk factors. J Magn Reson Imaging 2014;40(2):407-412.
23. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. Magn Reson Med 2004;52(1):141-146.
24. Piechnik SK, Ferreira VM, Dall'Armellina E, Cochlin LE, Greiser A, Neubauer S, Robson MD. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. J Cardiovasc Magn Reson 2010;12:69.
25. Radenkovic D, Weingartner S, Ricketts L, Moon JC, Captur G. T1 mapping in cardiac MRI. Heart Fail Rev 2017;22(4):415-430.
26. Higgins DM, Moon JC. Review of T1 Mapping Methods: Comparative Effectiveness Including Reproducibility Issues. Current Cardiovascular Imaging Reports 2014;7(3):9252.
27. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, Guimaraes AR, Filice R, Zhang Z, Sullivan DC, Group QTW. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 2015;24(1):9-26.

28. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, Kondratovich M, McShane LM, Reeves AP, Barboriak DP, Guimaraes AR, Wahl RL, Group R-QMW. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015;277(3):813-825.
29. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gonen M, Zahlmann G, Kondratovich MV, O'Donnell K, Petrick N, Cole PE, Garra B, Sullivan DC, Group QTPW. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27-67.
30. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, Barnhart HX, Jackson EF, Giger ML, Pennello G, Toledano AY, Kalpathy-Cramer J, Apanasovich TV, Kinahan PE, Myers KJ, Goldgof DB, Barboriak DP, Gillies RJ, Schwartz LH, Sullivan DC, Algorithm Comparison Working G. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. Stat Methods Med Res 2015;24(1):68-106.
31. Keenan KE, Ainslie M, Barker AJ, Boss MA, Cecil KM, Charles C, Chenevert TL, Clarke L, Evelhoch JL, Finn P, Gembris D, Gunter JL, Hill DLG, Jack CR, Jr., Jackson EF, Liu G, Russek SE, Sharma SD, Steckner M, Stupic KF, Trzasko JD, Yuan C, Zheng J. Quantitative magnetic resonance imaging phantoms: A review and the need for a system phantom. Magn Reson Med 2018;79(1):48-61.
32. Boss M, Dienstfrey A, Gimbutas Z, Keenan K, Splett J, Stupic K, Russek S. Magnetic Resonance Imaging Biomarker Calibration Service: Proton Spin Relaxation Times. Gaithersburg, MD: National Institute of Standards and Technology; 2018.
33. Stupic KF, Ainslie M, Boss MA, Charles C, Dienstfrey AM, Evelhoch JL, Finn P, Gimbutas Z, Gunter JL, Hill DLG, Jack CR, Jackson EF, Karaulanov T, Keenan KE, Liu G, Martin MN, Prasad PV, Rentz NS, Yuan C, Russek SE. A standard system phantom for magnetic resonance imaging. Magn Reson Med 2021.
34. Marques JP, Meineke J, Milovic C, Bilgic B, Chan KS, Hedouin R, van der Zwaag W, Langkammer C, Schweser F. QSM reconstruction challenge 2.0: A realistic in silico head phantom for MRI data simulation and evaluation of susceptibility mapping procedures. Magn Reson Med 2021;86(1):526-542.
35. Obuchowski NA, Buckler A, Kinahan P, Chen-Mayer H, Petrick N, Barboriak DP, Bullen J, Barnhart H, Sullivan DC. Statistical Issues in Testing Conformance with the Quantitative Imaging Biomarker Alliance (QIBA) Profile Claims. Acad Radiol 2016;23(4):496-506.
36. Obuchowski NA, Bullen J. Quantitative imaging biomarkers: effect of sample size and bias on confidence interval coverage. Stat Methods Med Res 2018;27(10):3139-3150.
37. Captur G, Bhandari A, Bruhl R, Ittermann B, Keenan KE, Yang Y, Eames RJ, Benedetti G, Torlasco C, Ricketts L, Boubertakh R, Fatih N, Greenwood JP, Paulis LEM, Lawton CB, Bucciarelli-Ducci C, Lamb HJ, Steeds R, Leung SW, Berry C, Valentin S, Flett A, de Lange C, DeCobelli F, Viallon M, Croisille P, Higgins DM, Greiser A, Pang W, Hamilton-Craig C, Strugnell WE, Dresselaers T, Barison A, Dawson D, Taylor AJ, Mongeon FP, Plein S, Messroghli D, Al-Mallah M, Grieve SM, Lombardi M, Jang J, Salerno M, Chaturvedi N, Kellman P, Bluemke DA, Nezafat R, Gatehouse P, Moon JC, Consortium TM. T1 mapping performance and measurement repeatability: results from

the multi-national T1 mapping standardization phantom program (T1MES). J Cardiovasc Magn Reson 2020;22(1):31.
38. Hernando D, Kuhn JP, Mensel B, Volzke H, Puls R, Hosten N, Reeder SB. R2* estimation using "in-phase" echoes in the presence of fat: the effects of complex spectrum of fat. J Magn Reson Imaging 2013;37(3):717-726.
39. Obuchowski NA, Mozley PD, Matthews D, Buckler A, Bullen J, Jackson E. Statistical Considerations for Planning Clinical Trials with Quantitative Imaging Biomarkers. J Natl Cancer Inst 2019;111(1):19-26.
40. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, Huang W, Noworolski SM, Young RJ, Shiroishi MS. 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(7):e101-e121.
41. Huang EP, Wang X-F, Choudhury KR, McShane LM, Gönen M, Ye J, Buckler AJ, Kinahan PE, Reeves AP, Jackson EF. Meta-analysis of the technical performance of an imaging procedure: guidelines and statistical methodology. Stat Methods Med Res 2015;24(1):141-174.
42. Obuchowski NA, Barnhart HX, Buckler AJ, Pennello G, Wang X-F, Kalpathy-Cramer J, Kim HJ, Reeves AP, Group CEW. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. Stat Methods Med Res 2015;24(1):107-140.
43. RSNA. QIBA Profile: Diffusion-Weighted Magnetic Resonance Imaging (DWI). 2019.
44. Hernando D, Cook RJ, Qazi N, Longhurst CA, Diamond CA, Reeder SB. Complex confounder-corrected R2* mapping for liver iron quantification with MRI. Eur Radiol 2021;31(1):264-275.
45. Hernando D, Zhao R, Mattison R, Reeder S, Yokoo T, Pedrosa I, Yuan Q, Karampinos D, Ruschke S, Kamel I, Aliyari M, Zhong X, Vasanawala S, Taviani V. Multi-center, multi-vendor reproducibility of confounder-corrected R2* mapping for liver iron quantification at 1.5T and 3T: interim results. 2019 March 17, 2019; Orlando, FL. Society of Abdominal Imaging Annual Meeting.
46. Hernando D, Sharma SD, Aliyari Ghasabeh M, Alvis BD, Arora SS, Hamilton G, Pan L, Shaffer JM, Sofue K, Szeverenyi NM, Welch EB, Yuan Q, Bashir MR, Kamel IR, Rice MJ, Sirlin CB, Yokoo T, Reeder SB. Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat-water phantom. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2017;77(4):1516-1524.
47. Hu HH, Yokoo T, Bashir MR, Sirlin CB, Hernando D, Malyarenko D, Chenevert TL, Smith MA, Serai SD, Middleton MS, Henderson WC, Hamilton G, Shaffer J, Shu Y, Tkach JA, Trout AT, Obuchowski N, Brittain JH, Jackson EF, Reeder SB, Committee RQIBAPB. Linearity and Bias of Proton Density Fat Fraction as a Quantitative Imaging Biomarker: A Multicenter, Multiplatform, Multivendor Phantom Study. Radiology 2021;298(3):640-651.
48. Yokoo T, Serai SD, Pirasteh A, Bashir MR, Hamilton G, Hernando D, Hu HH, Hetterich H, Kuhn JP, Kukuk GM, Loomba R, Middleton MS, Obuchowski NA, Song JS, Tang A, Wu X, Reeder SB, Sirlin CB, Committee R-QPB. Linearity, Bias, and Precision of

Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis. Radiology 2018;286(2):486-498.
49. Bannas P, Kramer H, Hernando D, Agni R, Cunningham AM, Mandal R, Motosugi U, Sharma SD, Munoz del Rio A, Fernandez L, Reeder SB. Quantitative magnetic resonance imaging of hepatic steatosis: Validation in ex vivo human livers. Hepatology 2015;62(5):1444-1455.
50. Hines CD, Frydrychowicz A, Hamilton G, Tudorascu DL, Vigen KK, Yu H, McKenzie CA, Sirlin CB, Brittain JH, Reeder SB. T(1) independent, T(2) (*) corrected chemical shift based fat-water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. Journal of magnetic resonance imaging : JMRI 2011;33(4):873-881.
51. Serai SD, Dillman JR, Trout AT. Proton Density Fat Fraction Measurements at 1.5- and 3-T Hepatic MR Imaging: Same-Day Agreement among Readers and across Two Imager Manufacturers. Radiology 2017;284(1):244-254.
52. Roujol S, Weingartner S, Foppa M, Chow K, Kawaji K, Ngo LH, Kellman P, Manning WJ, Thompson RB, Nezafat R. Accuracy, precision, and reproducibility of four T1 mapping sequences: a head-to-head comparison of MOLLI, ShMOLLI, SASHA, and SAPPHIRE. Radiology 2014;272(3):683-689.
53. Weingartner S, Messner NM, Budjan J, Lossnitzer D, Mattler U, Papavassiliu T, Zollner FG, Schad LR. Myocardial T1-mapping at 3T using saturation-recovery: reference values, precision and comparison with MOLLI. J Cardiovasc Magn Reson 2016;18(1):84.
54. Fontana M, White SK, Banypersad SM, Sado DM, Maestrini V, Flett AS, Piechnik SK, Neubauer S, Roberts N, Moon JC. Comparison of T1 mapping techniques for ECV quantification. Histological validation and reproducibility of ShMOLLI versus multibreath-hold T1 quantification equilibrium contrast CMR. J Cardiovasc Magn Reson 2012;14:88.
55. Kellman P, Hansen MS. T1-mapping in the heart: accuracy and precision. J Cardiovasc Magn Reson 2014;16:2.
56. Graham-Brown MP, Rutherford E, Levelt E, March DS, Churchward DR, Stensel DJ, McComb C, Mangion K, Cockburn S, Berry C, Moon JC, Mark PB, Burton JO, McCann GP. Native T1 mapping: inter-study, inter-observer and inter-center reproducibility in hemodialysis patients. J Cardiovasc Magn Reson 2017;19(1):21.
57. Chow K, Flewitt JA, Green JD, Pagano JJ, Friedrich MG, Thompson RB. Saturation recovery single-shot acquisition (SASHA) for myocardial T(1) mapping. Magn Reson Med 2014;71(6):2082-2095.
58. Raman FS, Kawel-Boehm N, Gai N, Freed M, Han J, Liu CY, Lima JA, Bluemke DA, Liu S. Modified look-locker inversion recovery T1 mapping indices: assessment of accuracy and reproducibility between magnetic resonance scanners. J Cardiovasc Magn Reson 2013;15:64.
59. Dabir D, Child N, Kalra A, Rogers T, Gebker R, Jabbour A, Plein S, Yu CY, Otton J, Kidambi A, McDiarmid A, Broadbent D, Higgins DM, Schnackenburg B, Foote L, Cummins C, Nagel E, Puntmann VO. Reference values for healthy human myocardium using a T1 mapping methodology: results from the International T1 Multicenter cardiovascular magnetic resonance study. J Cardiovasc Magn Reson 2014;16:69.

60. Popescu IA, Werys K, Zhang Q, Puchta H, Hann E, Lukaschuk E, Ferreira VM, Piechnik SK. Standardization of T1-mapping in cardiovascular magnetic resonance using clustered structuring for benchmarking normal ranges. Int J Cardiol 2021;326:220-225.
61. Glover GH, Schneider E. Three-point Dixon technique for true water/fat decomposition with B0 inhomogeneity correction. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 1991;18(2):371-383.
62. Malyarenko DI, Ross BD, Chenevert TL. Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements. Magn Reson Med 2014;71(3):1312-1323.
63. Bernstein MA, Zhou XJ, Polzin JA, King KF, Ganin A, Pelc NJ, Glover GH. Concomitant gradient terms in phase contrast MR: analysis and correction. Magn Reson Med 1998;39(2):300-308.
64. Colgan TJ, Hernando D, Sharma SD, Reeder SB. The effects of concomitant gradients on chemical shift encoded MRI. Magn Reson Med 2017;78(2):730-738.
65. Jezzard P, Barnett AS, Pierpaoli C. Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. Magn Reson Med 1998;39(5):801-812.
66. Benner T, van der Kouwe AJ, Kirsch JE, Sorensen AG. Real-time RF pulse adjustment for B0 drift correction. Magn Reson Med 2006;56(1):204-209.
67. Peters DC, Derbyshire JA, McVeigh ER. Centering the projection reconstruction trajectory: reducing gradient delay errors. Magn Reson Med 2003;50(1):1-6.
68. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: A complex problem with many partial solutions. J Magn Reson Imaging 2015;42(4):887-901.
69. Novikov DS, Kiselev VG, Jespersen SN. On modeling. Magn Reson Med 2018;79(6):3172-3193.
70. McRobbie D, Lerski R, Straughan K. Slice profile effects and their calibration and correction in quantitative NMR imaging. Physics in Medicine & Biology 1987;32(8):971.
71. Malik SJ, Kenny GD, Hajnal JV. Slice profile correction for transmit sensitivity mapping using actual flip angle imaging. Magn Reson Med 2011;65(5):1393-1399.
72. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 1995;34(6):910-914.
73. Jones DK, Basser PJ. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn Reson Med 2004;52(5):979-993.
74. Hernando D, Kramer JH, Reeder SB. Multipeak fat-corrected complex R2* relaxometry: theory, optimization, and clinical validation. Magn Reson Med 2013;70(5):1319-1331.
75. Bydder M, Yokoo T, Hamilton G, Middleton MS, Chavez AD, Schwimmer JB, Lavine JE, Sirlin CB. Relaxation effects in the quantification of fat using gradient echo imaging. Magnetic resonance imaging 2008;26(3):347-359.
76. Horng DE, Hernando D, Reeder SB. Quantification of liver fat in the presence of iron overload. J Magn Reson Imaging 2017;45(2):428-439.
77. Hamilton G, Yokoo T, Bydder M, Cruite I, Schroeder ME, Sirlin CB, Middleton MS. In vivo characterization of the liver fat (1)H MR spectrum. NMR in biomedicine 2011;24(7):784-790.
78. Yu H, Shimakawa A, Hines CD, McKenzie CA, Hamilton G, Sirlin CB, Brittain JH, Reeder SB. Combination of complex-based and magnitude-based multiecho water-fat

separation for accurate quantification of fat-fraction. Magn Reson Med 2011;66(1):199-206.
79. Hernando D, Hines CD, Yu H, Reeder SB. Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2012;67(3):638-644.
80. Weingartner S, Akcakaya M, Basha T, Kissinger KV, Goddu B, Berg S, Manning WJ, Nezafat R. Combined saturation/inversion recovery sequences for improved evaluation of scar and diffuse fibrosis in patients with arrhythmia or heart rate variability. Magn Reson Med 2014;71(3):1024-1034.
81. Christodoulou AG, Shaw JL, Nguyen C, Yang Q, Xie Y, Wang N, Li D. Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging. Nat Biomed Eng 2018;2(4):215-226.
82. Weingartner S, Shenoy C, Rieger B, Schad LR, Schulz-Menger J, Akcakaya M. Temporally resolved parametric assessment of Z-magnetization recovery (TOPAZ): Dynamic myocardial T1 mapping using a cine steady-state look-locker approach. Magn Reson Med 2018;79(4):2087-2100.
83. Guo R, Cai X, Kucukseymen S, Rodriguez J, Paskavitz A, Pierce P, Goddu B, Thompson RB, Nezafat R. Free-breathing simultaneous myocardial T1 and T2 mapping with whole left ventricle coverage. Magn Reson Med 2021;85(3):1308-1321.
84. Robson MD, Piechnik SK, Tunnicliffe EM, Neubauer S. T1 measurements in the human myocardium: the effects of magnetization transfer on the SASHA and MOLLI sequences. Magn Reson Med 2013;70(3):664-670.
85. Kellman P, Herzka DA, Arai AE, Hansen MS. Influence of Off-resonance in myocardial T1-mapping using SSFP based MOLLI method. J Cardiovasc Magn Reson 2013;15:63.
86. Cooper MA, Nguyen TD, Spincemaille P, Prince MR, Weinsaft JW, Wang Y. How Accurate Is MOLLI T1 Mapping In Vivo? Validation by Spin Echo Methods. PLOS ONE 2014;9(9):e107327.
87. Cameron D, Vassiliou VS, Higgins DM, Gatehouse PD. Towards accurate and precise T 1 and extracellular volume mapping in the myocardium: a guide to current pitfalls and their solutions. MAGMA 2018;31(1):143-163.
88. Weingartner S, Messner NM, Zollner FG, Akcakaya M, Schad LR. Black-blood native T1 mapping: Blood signal suppression for reduced partial voluming in the myocardium. Magn Reson Med 2017;78(2):484-493.
89. Motosugi U, Hernando D, Wiens C, Bannas P, Reeder SB. High SNR Acquisitions Improve the Repeatability of Liver Fat Quantification Using Confounder-corrected Chemical Shift-encoded MR Imaging. Magn Reson Med Sci 2017;16(4):332-339.
90. Boudreau M, Pike GB. Sensitivity regularization of the Cramér-Rao lower bound to minimize B1 nonuniformity effects in quantitative magnetization transfer imaging. Magn Reson Med 2018;80(6):2560-2572.
91. Drakesmith M, Harms R, Rudrapatna SU, Parker GD, Evans CJ, Jones DK. Estimating axon conduction velocity in vivo from microstructural MRI. Neuroimage 2019;203:116186.
92. Scharf LL, LT M. Geometry of the Cramér-Rao Bound. In: Proceedings of the IEEE Sixth SP Workshop on Statistical Signal and Array Processing. Victoria, BC, Canada1992. p 5-8.

93. Akcakaya M, Weingartner S, Roujol S, Nezafat R. On the selection of sampling points for myocardial T1 mapping. Magn Reson Med 2015;73(5):1741-1753.
94. Kellman P, Xue H, Chow K, Spottiswoode BS, Arai AE, Thompson RB. Optimized saturation recovery protocols for T1-mapping in the heart: influence of sampling strategies on precision. J Cardiovasc Magn Reson 2014;16:55.
95. Jang J, Bellm S, Roujol S, Basha TA, Nezafat M, Kato S, Weingartner S, Nezafat R. Comparison of spoiled gradient echo and steady-state free-precession imaging for native myocardial T1 mapping using the slice-interleaved T1 mapping (STONE) sequence. NMR Biomed 2016;29(10):1486-1496.
96. Shaw JL, Yang Q, Zhou Z, Deng Z, Nguyen C, Li D, Christodoulou AG. Free-breathing, non-ECG, continuous myocardial T1 mapping with cardiovascular magnetic resonance multitasking. Magn Reson Med 2019;81(4):2450-2463.
97. Qi H, Jaubert O, Bustin A, Cruz G, Chen H, Botnar R, Prieto C. Free-running 3D whole heart myocardial T1 mapping with isotropic spatial resolution. Magn Reson Med 2019;82(4):1331-1342.
98. Chow K, Yang Y, Shaw P, Kramer CM, Salerno M. Robust free-breathing SASHA T1 mapping with high-contrast image registration. J Cardiovasc Magn Reson 2016;18(1):47.
99. Weingartner S, Roujol S, Akcakaya M, Basha TA, Nezafat R. Free-breathing multislice native myocardial T1 mapping using the slice-interleaved T1 (STONE) sequence. Magn Reson Med 2015;74(1):115-124.
100. Weingartner S, Akcakaya M, Roujol S, Basha T, Stehning C, Kissinger KV, Goddu B, Berg S, Manning WJ, Nezafat R. Free-breathing post-contrast three-dimensional T1 mapping: Volumetric assessment of myocardial T1 values. Magn Reson Med 2015;73(1):214-222.
101. Nordio G, Bustin A, Henningsson M, Rashid I, Chiribiri A, Ismail T, Odille F, Prieto C, Botnar RM. 3D SASHA myocardial T1 mapping with high accuracy and improved precision. MAGMA 2019;32(2):281-289.
102. Wang Y, Liu T. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. Magn Reson Med 2015;73(1):82-101.
103. Hancu I, Liu J, Hua Y, Lee SK. Electrical properties tomography: Available contrast and reconstruction capabilities. Magn Reson Med 2019;81(2):803-810.
104. Liu J, Wang Y, Katscher U, He B. Electrical Properties Tomography Based on $B_{{1}}$ Maps in MRI: Principles, Applications, and Challenges. IEEE Trans Biomed Eng 2017;64(11):2515-2530.
105. Bydder M, Yokoo T, Yu H, Carl M, Reeder SB, Sirlin CB. Constraining the initial phase in water-fat separation. Magnetic resonance imaging 2011;29(2):216-221.
106. Horng DE, Hernando D, Hines CD, Reeder SB. Comparison of R2* correction methods for accurate fat quantification in fatty liver. J Magn Reson Imaging 2013;37(2):414-422.
107. Wang X, Hernando D, Reeder SB. Sensitivity of chemical shift-encoded fat quantification to calibration of fat MR spectrum. Magn Reson Med 2016;75(2):845-851.
108. Deichmann R, Haase A. Quantification of T1 values by SNAPSHOT-FLASH NMR imaging. Journal of Magnetic Resonance (1969) 1992;96(3):608-612.
109. Shao J, Rapacchi S, Nguyen KL, Hu P. Myocardial T1 mapping at 3.0 tesla using an inversion recovery spoiled gradient echo readout and bloch equation simulation with slice profile correction (BLESSPC) T1 estimation algorithm. J Magn Reson Imaging 2016;43(2):414-425.

110. Xanthis CG, Bidhult S, Kantasis G, Heiberg E, Arheden H, Aletras AH. Parallel simulations for QUAntifying RElaxation magnetic resonance constants (SQUAREMR): an example towards accurate MOLLI T1 measurements. J Cardiovasc Magn Reson 2015;17:104.
111. Veraart J, Sijbers J, Sunaert S, Leemans A, Jeurissen B. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. Neuroimage 2013;81:335-346.
112. Bertsekas DP. Nonlinear Programming: Athena Scientific; 2016.
113. Xue H, Greiser A, Zuehlsdorff S, Jolly MP, Guehring J, Arai AE, Kellman P. Phase-sensitive inversion recovery for myocardial T1 mapping with motion correction and parametric fitting. Magn Reson Med 2013;69(5):1408-1420.
114. Hunter DJ, Losina E, Guermazi A, Burstein D, Lassere MN, Kraus V. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. Curr Drug Targets 2010;11(5):536-545.
115. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood) 2018;243(3):213-221.
116. Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Abramson RG, Burton KR, John-Paul JY, Scalzetti EM, Yankeelov TE, Subramaniam RM. Clinical utility of quantitative imaging. Acad Radiol 2015;22(1):33-49.
117. Brisman JL, Pile-Spellman J, Konstas AA. Clinical utility of quantitative magnetic resonance angiography in the assessment of the underlying pathophysiology in a variety of cerebrovascular disorders. Eur J Radiol 2012;81(2):298-302.
118. Eskreis-Winkler S, Zhang Y, Zhang J, Liu Z, Dimov A, Gupta A, Wang Y. The clinical utility of QSM: disease diagnosis, medical management, and surgical planning. NMR Biomed 2017;30(4):e3668.
119. Gallagher EJ. Clinical utility of likelihood ratios. Ann Emerg Med 1998;31(3):391-397.
120. Dobbin KK, Cesano A, Alvarez J, Hawtin R, Janetzki S, Kirsch I, Masucci GV, Robbins PB, Selvan SR, Streicher HZ, Zhang J, Butterfield LH, Thurin M. Validation of biomarkers to predict response to immunotherapy in cancer: Volume II - clinical validation and regulatory considerations. J Immunother Cancer 2016;4:77.
121. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins D, Hammoud DA, Rustin GJ, Taouli B, Choyke PL. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 2009;11(2):102-125.
122. Baessler B, Schaarschmidt F, Dick A, Stehning C, Schnackenburg B, Michels G, Maintz D, Bunck AC. Mapping tissue inhomogeneity in acute myocarditis: a novel analytical approach to quantitative myocardial edema imaging by T2-mapping. J Cardiovasc Magn Reson 2015;17:115.
123. Baessler B, Schaarschmidt F, Stehning C, Schnackenburg B, Giolda A, Maintz D, Bunck AC. Reproducibility of three different cardiac T2 -mapping sequences at 1.5T. J Magn Reson Imaging 2016;44(5):1168-1178.
124. Gu J, Liu S, Du S, Zhang Q, Xiao J, Dong Q, Xin Y. Diagnostic value of MRI-PDFF for hepatic steatosis in patients with non-alcoholic fatty liver disease: a meta-analysis. Eur Radiol 2019;29(7):3564-3573.
125. Loomba R, Neuschwander-Tetri BA, Sanyal A, Chalasani N, Diehl AM, Terrault N, Kowdley K, Dasarathy S, Kleiner D, Behling C, Lavine J, Van Natta M, Middleton M,

Tonascia J, Sirlin C, Network NCR. Multicenter Validation of Association Between Decline in MRI-PDFF and Histologic Response in NASH. Hepatology 2020;72(4):1219-1229.
126. Stine JG, Munaganuru N, Barnard A, Wang JL, Kaulback K, Argo CK, Singh S, Fowler KJ, Sirlin CB, Loomba R. Change in MRI-PDFF and Histologic Response in Patients With Nonalcoholic Steatohepatitis: A Systematic Review and Meta-Analysis. Clin Gastroenterol Hepatol 2020.
127. Rehm JL, Wolfgram PM, Hernando D, Eickhoff JC, Allen DB, Reeder SB. Proton density fat-fraction is an accurate biomarker of hepatic steatosis in adolescent girls and young women. Eur Radiol 2015;25(10):2921-2930.
128. Puchner SB, Lu MT, Mayrhofer T, Liu T, Pursnani A, Ghoshhajra BB, Truong QA, Wiviott SD, Fleg JL, Hoffmann U, Ferencik M. High-risk coronary plaque at coronary CT angiography is associated with nonalcoholic fatty liver disease, independent of coronary plaque and stenosis burden: results from the ROMICAT II trial. Radiology 2015;274(3):693-701.
129. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67(1):328-357.
130. Faragli A, Merz S, Muzio FPL, Doeblin P, Tanacli R, Kolp C, Abawi D, Otvos J, Stehning C, Schnackenburg B, Pieske B, Post H, Klopfleisch R, Alogna A, Kelle S. Estimation of total collagen volume: a T1 mapping versus histological comparison study in healthy Landrace pigs. Int J Cardiovasc Imaging 2020;36(9):1761-1769.
131. Reiter U, Reiter C, Krauter C, Fuchsjager M, Reiter G. Cardiac magnetic resonance T1 mapping. Part 2: Diagnostic potential and applications. Eur J Radiol 2018;109:235-247.
132. Hamilton-Craig CR, Strudwick MW, Galloway GJ. T1 Mapping for Myocardial Fibrosis by Cardiac Magnetic Resonance Relaxometry-A Comprehensive Technical Review. Front Cardiovasc Med 2016;3:49.
133. Lurz JA, Luecke C, Lang D, Besler C, Rommel KP, Klingel K, Kandolf R, Adams V, Schöne K, Hindricks G, Schuler G, Linke A, Thiele H, Gutberlet M, Lurz P. CMR-Derived Extracellular Volume Fraction as a Marker for Myocardial Fibrosis: The Importance of Coexisting Myocardial Inflammation. JACC Cardiovasc Imaging 2018;11(1):38-45.
134. Fehrmann A, Treutlein M, Rudolph T, Rudolph V, Weiss K, Giese D, Bunck AC, Maintz D, Baessler B. Myocardial T1 and T2 mapping in severe aortic stenosis: Potential novel insights into the pathophysiology of myocardial remodelling. Eur J Radiol 2018;107:76-83.
135. Aherne E, Chow K, Carr J. Cardiac T1 mapping: Techniques and applications. J Magn Reson Imaging 2020;51(5):1336-1356.
136. Pan JA, Kerwin MJ, Salerno M. Native T1 Mapping, Extracellular Volume Mapping, and Late Gadolinium Enhancement in Cardiac Amyloidosis: A Meta-Analysis. JACC Cardiovasc Imaging 2020;13(6):1299-1310.
137. Duan C, Zhu Y, Jang J, Rodriguez J, Neisius U, Fahmy AS, Nezafat R. Non-contrast myocardial infarct scar assessment using a hybrid native T1 and magnetization transfer imaging sequence at 1.5T. Magn Reson Med 2019;81(5):3192-3201.

138. Lopez K, Neji R, Mukherjee RK, Whitaker J, Phinikaridou A, Razavi R, Prieto C, Roujol S, Botnar R. Contrast-free high-resolution 3D magnetization transfer imaging for simultaneous myocardial scar and cardiac vein visualization. MAGMA 2020;33(5):627-640.
139. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Faivre-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh DM, Lacombe D, Lambin P, Lassau N, Leach MO, Lee TY, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJ, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Stroobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14(3):169-186.
140. Tofts PS, Collins DJ. Multicentre imaging measurements for oncology and in the brain. Br J Radiol 2011;84 Spec No 2(Spec Iss 2):S213-226.
141. Captur G, Gatehouse P, Keenan KE, Heslinga FG, Bruehl R, Prothmann M, Graves MJ, Eames RJ, Torlasco C, Benedetti G, Donovan J, Ittermann B, Boubertakh R, Bathgate A, Royet C, Pang W, Nezafat R, Salerno M, Kellman P, Moon JC. A medical device-grade T1 and ECV phantom for global T1 mapping quality assurance-the T1 Mapping and ECV Standardization in cardiovascular magnetic resonance (T1MES) program. J Cardiovasc Magn Reson 2016;18(1):58.
142. Keenan KE, Biller JR, Delfino JG, Boss MA, Does MD, Evelhoch JL, Griswold MA, Gunter JL, Hinks RS, Hoffman SW, Kim G, Lattanzi R, Li X, Marinelli L, Metzger GJ, Mukherjee P, Nordstrom RJ, Peskin AP, Perez E, Russek SE, Sahiner B, Serkova N, Shukla-Dave A, Steckner M, Stupic KF, Wilmes LJ, Wu HH, Zhang H, Jackson EF, Sullivan DC. Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. J Magn Reson Imaging 2019;49(7):e26-e39.
143. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Aerts HJ, Bendriem B, Bendtsen C, Boellaard R, Boone JM, Cole PE, Conklin JJ, Dorfman GS, Douglas PS, Eidsaunet W, Elsinger C, Frank RA, Gatsonis C, Giger ML, Gupta SN, Gustafson D, Hoekstra OS, Jackson EF, Karam L, Kelloff GJ, Kinahan PE, McLennan G, Miller CG, Mozley PD, Muller KE, Patt R, Raunig D, Rosen M, Rupani H, Schwartz LH, Siegel BA, Sorensen AG, Wahl RL, Waterton JC, Wolf W, Zahlmann G, Zimmerman B. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 2011;259(3):875-884.
144. Radiological Society of North America Quantitative Imaging Biomarkers Alliance.
145. Kuhn JP, Meffert P, Heske C, Kromrey ML, Schmidt CO, Mensel B, Volzke H, Lerch MM, Hernando D, Mayerle J, Reeder SB. Prevalence of Fatty Liver Disease and Hepatic Iron Overload in a Northeastern German Population by Using Quantitative MR Imaging. Radiology 2017;284(3):706-716.
146. Ajmera V, Park CC, Caussy C, Singh S, Hernandez C, Bettencourt R, Hooker J, Sy E, Behling C, Xu R, Middleton MS, Valasek MA, Faulkner C, Rizo E, Richards L, Sirlin CB, Loomba R. Magnetic Resonance Imaging Proton Density Fat Fraction Associates

With Progression of Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2018;155(2):307-310 e302.
147. Baggiano A, Boldrini M, Martinez-Naharro A, Kotecha T, Petrie A, Rezk T, Gritti M, Quarta C, Knight DS, Wechalekar AD, Lachmann HJ, Perlini S, Pontone G, Moon JC, Kellman P, Gillmore JD, Hawkins PN, Fontana M. Noncontrast Magnetic Resonance for the Diagnosis of Cardiac Amyloidosis. JACC Cardiovasc Imaging 2020;13(1 Pt 1):69-80.
148. Augusto JB, Nordin S, Vijapurapu R, Baig S, Bulluck H, Castelletti S, Alfarih M, Knott K, Captur G, Kotecha T, Ramaswami U, Tchan M, Geberhiwot T, Fontana M, Steeds RP, Hughes D, Kozor R, Moon JC. Myocardial Edema, Myocyte Injury, and Disease Severity in Fabry Disease. Circ Cardiovasc Imaging 2020;13(3):e010171.
149. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, Young AA, Hudson S, Weale P, Garratt S, Collins R, Piechnik S, Neubauer S. UK Biobank's cardiovascular magnetic resonance protocol. J Cardiovasc Magn Reson 2016;18:8.
150. Bamberg F, Kauczor HU, Weckbach S, Schlett CL, Forsting M, Ladd SC, Greiser KH, Weber MA, Schulz-Menger J, Niendorf T, Pischon T, Caspers S, Amunts K, Berger K, Bulow R, Hosten N, Hegenscheid K, Kroncke T, Linseisen J, Gunther M, Hirsch JG, Kohn A, Hendel T, Wichmann HE, Schmidt B, Jockel KH, Hoffmann W, Kaaks R, Reiser MF, Volzke H, German National Cohort MRISI. Whole-Body MR Imaging in the German National Cohort: Rationale, Design, and Technical Background. Radiology 2015;277(1):206-220.
151. Ferreira VM, Schulz-Menger J, Holmvang G, Kramer CM, Carbone I, Sechtem U, Kindermann I, Gutberlet M, Cooper LT, Liu P, Friedrich MG. Cardiovascular Magnetic Resonance in Nonischemic Myocardial Inflammation: Expert Recommendations. J Am Coll Cardiol 2018;72(24):3158-3176.
152. van Beek EJR, Kuhl C, Anzai Y, Desmond P, Ehman RL, Gong Q, Gold G, Gulani V, Hall-Craggs M, Leiner T, Lim CCT, Pipe JG, Reeder S, Reinhold C, Smits M, Sodickson DK, Tempany C, Vargas HA, Wang M. Value of MRI in medicine: More than just another test? J Magn Reson Imaging 2019;49(7):e14-e25.
153. Farahani K, Tata D, Nordstrom RJ. QIN Benchmarks for Clinical Translation of Quantitative Imaging Tools. Tomography 2019;5(1):1-6.
154. Wilson M, Andronesi O, Barker PB, Bartha R, Bizzi A, Bolan PJ, Brindle KM, Choi IY, Cudalbu C, Dydak U, Emir UE, Gonzalez RG, Gruber S, Gruetter R, Gupta RK, Heerschap A, Henning A, Hetherington HP, Huppi PS, Hurd RE, Kantarci K, Kauppinen RA, Klomp DWJ, Kreis R, Kruiskamp MJ, Leach MO, Lin AP, Luijten PR, Marjanska M, Maudsley AA, Meyerhoff DJ, Mountford CE, Mullins PG, Murdoch JB, Nelson SJ, Noeske R, Oz G, Pan JW, Peet AC, Poptani H, Posse S, Ratai EM, Salibi N, Scheenen TWJ, Smith ICP, Soher BJ, Tkac I, Vigneron DB, Howe FA. Methodological consensus on clinical proton MRS of the brain: Review and recommendations. Magn Reson Med 2019;82(2):527-550.
155. Near J, Harris AD, Juchem C, Kreis R, Marjanska M, Oz G, Slotboom J, Wilson M, Gasparovic C. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. NMR Biomed 2021;34(5):e4257.
156. Alsop DC, Detre JA, Golay X, Gunther M, Hendrikse J, Hernandez-Garcia L, Lu H, MacIntosh BJ, Parkes LM, Smits M, van Osch MJ, Wang DJ, Wong EC, Zaharchuk G. Recommended implementation of arterial spin-labeled perfusion MRI for clinical

applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med 2015;73(1):102-116.
157. Taouli B, Beer AJ, Chenevert T, Collins D, Lehman C, Matos C, Padhani AR, Rosenkrantz AB, Shukla-Dave A, Sigmund E, Tanenbaum L, Thoeny H, Thomassin-Naggara I, Barbieri S, Corcuera-Solano I, Orton M, Partridge SC, Koh DM. Diffusion-weighted imaging outside the brain: Consensus statement from an ISMRM-sponsored workshop. J Magn Reson Imaging 2016;44(3):521-540.
158. Ljimani A, Caroli A, Laustsen C, Francis S, Mendichovszky IA, Bane O, Nery F, Sharma K, Pohlmann A, Dekkers IA, Vallee JP, Derlin K, Notohamiprodjo M, Lim RP, Palmucci S, Serai SD, Periquito J, Wang ZJ, Froeling M, Thoeny HC, Prasad P, Schneider M, Niendorf T, Pullens P, Sourbron S, Sigmund EE. Correction to: Consensus-based technical recommendations for clinical translation of renal diffusion-weighted MRI. MAGMA 2020;33(1):197-198.
159. Ljimani A, Caroli A, Laustsen C, Francis S, Mendichovszky IA, Bane O, Nery F, Sharma K, Pohlmann A, Dekkers IA, Vallee JP, Derlin K, Notohamiprodjo M, Lim RP, Palmucci S, Serai SD, Periquito J, Wang ZJ, Froeling M, Thoeny HC, Prasad P, Schneider M, Niendorf T, Pullens P, Sourbron S, Sigmund EE. Consensus-based technical recommendations for clinical translation of renal diffusion-weighted MRI. MAGMA 2020;33(1):177-195.
160. Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, Martincich L, Partridge SC, Patterson A, Pinker K, Thibault F, Camps-Herrero J, Le Bihan D, group EiBD-WIw. 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):1436-1450.
161. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, Huang W, Noworolski SM, Young RJ, Shiroishi MS, Kim H, Coolens C, Laue H, Chung C, Rosen M, Boss M, Jackson EF. 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(7):e101-e121.
162. Dyverfeldt P, Bissell M, Barker AJ, Bolger AF, Carlhall CJ, Ebbers T, Francios CJ, Frydrychowicz A, Geiger J, Giese D, Hope MD, Kilner PJ, Kozerke S, Myerson S, Neubauer S, Wieben O, Markl M. 4D flow cardiovascular magnetic resonance consensus statement. J Cardiovasc Magn Reson 2015;17:72.
163. Committee QSMCO, Bilgic B, Langkammer C, Marques JP, Meineke J, Milovic C, Schweser F. QSM reconstruction challenge 2.0: Design and report of results. Magn Reson Med 2021;86(3):1241-1255.
164. Karakuzu A, Boudreau M, Duval T, Boshkovski T, Leppert I, Cabana J, Gagnon I, Beliveau P, Pike G, Cohen-Adad J, Stikov N. qMRLab: Quantitative MRI analysis, under one umbrella. Journal of Open Source Software 2020;5(53):2343.
165. Ma D, Gulani V, Seiberlich N, Liu K, Sunshine JL, Duerk JL, Griswold MA. Magnetic resonance fingerprinting. Nature 2013;495(7440):187-192.

Figure Captions:

Figure 1: Example quantitative MR methods illustrated in this manuscript. (Top) Liver proton-density fat fraction (PDFF) mapping, with applications in the evaluation of non-alcoholic fatty liver disease (where liver PDFF is an emerging biomarker for the early diagnosis of NAFLD) and non-alcoholic steatohepatitis (where liver PDFF is emerging as a medical research tool in combination with other noninvasive imaging biomarkers). (Bottom) Cardiac T_1 mapping, with applications in various ischemic and non-ischemic cardiomyopathies.

Figure 2: The major performance metrics for quantitative MR methods are the bias and precision profiles. Bias measures the systematic differences between the measurements and the ground truth. Evaluation of bias requires a highly reliable reference method. Precision measures the tendency of the qMR method to produce different values, when applied repeatedly on the same subject. Evaluation of precision does not require reference values. The term 'profile' refers to the potential variability of performance (bias or precision) across different true values. (SD: standard deviation; CV: coefficient of variation).

Figure 3. Effect of imprecision on the minimum detectable change for a quantitative method. As the within-subject standard deviation (wSD) increases, the minimum detectable change also increases. In this plot, both are given in the units of the measurand. For PDFF, if subjects were always imaged with the same imaging system, we might expect wSD~1%, resulting in a minimum detectable change (95% confidence) of less than 3% absolute PDFF. If subjects were allowed to be imaged with a variety of imaging systems, we might expect wSD closer to 2%, resulting in a minimum detectable change of more than 5% absolute PDFF.

Figure 4: R_2^* decay, if uncorrected, can confound PDFF quantification, leading to bias, as well as poor precision (eg: poor reproducibility across acquisitions with different number of echoes), particularly in patients with elevated liver $R_2^*=1/T_2^*$ ($R_2^*=160$ s^{-1} at 1.5T, corresponding to mild iron overload). As shown though simulation and in vivo, R_2^*-uncorrected signal fitting results are highly dependent on the choice of echo times. In contrast, R_2^*-corrected PDFF quantification has low bias and high reproducibility across choices of echo times. For this illustration, a 12-echo liver CSE acquisition in a patient with high liver fat and iron overload was reprocessed

retrospectively multiple times, using the first n echoes (for n=5,…,12). In each case, both R_2^*-uncorrected and R_2^*-corrected PDFF mapping methods were used.

Figure 5: Technical development and validation is typically an iterative process including technical implementation and refinements, as well as evaluation of bias and/or precision. The later stages of the process are typically more costly, often focused on a specific organ, patient population, and application (context of use), and often include multi-center validation. Once development and validation are completed for one application, extension to other applications is often of interest (e.g., PDFF measurements in skeletal muscle or bone marrow). This extension typically requires additional refinement/validation iterations. Beyond the technical development and validation described in this figure, subsequent clinical qualification (not shown) is needed to establish the relationship between the qMR measurement and specific biological processes or clinical endpoints of interest.

Figure 6: Steps and challenges for the validation, qualification, and dissemination of quantitative MR methods. Technical development and validation as well as clinical qualification are needed in order to establish the performance and clinical utility of qMR methods. Even with successful validation and qualification, substantial challenges (including regulatory and market-based factors) need to be overcome in order to achieve widespread dissemination. Importantly, these three processes can be advanced in parallel as suggested by the horizontal overlap. For example, technical development and validation can be performed at the same time as clinical qualification.

Figure 7: Schematic representation to illustrate that biased measurements can lead to artificially inflated effect size and discrimination between healthy and diseased states, if the confounders of the measurement happen to be sensitive to the pathological alteration. In the example of myocardial T_1 mapping, the most widely used mapping technique MOLLI is known to be confounded by several factors including T_2 time and magnetization transfer (MT) of the tissue. For a specific disease, however, this bias may accentuate the difference between healthy and disease and lead to larger effect sizes compared with unbiased T_1 quantification. However, this gain in effect size comes at the cost of reduced reproducibility as the factors contributing to the bias may vary across acquisitions, systems, and patients.

Supporting Information Table and Figure Captions:

Commented [A50]: E.3

Supporting Information Table S1: Glossary.

Supporting Information Table S2: Individuals and organizations endorsing this manuscript.

Supporting Information Figure S1. Question 1: In what area are the qMR metrics that you study?

Supporting Information Figure S2. Question 2: Please select the confounders you encounter and the frequency with which they appear.

Supporting Information Figure S3. Question 3: What steps do you take to correct confounders, and how often are they implemented in your qMR protocols?

Supporting Information Figure S4. Question 4: Many qMR metrics are promising during the development stage but fail to cross translational gaps. What are the most common hurdles that you face between successful technical development and clinical application? Please evaluate the following in terms of resources required to overcome these hurdles.

Supporting Information Figure S5. Question 5: What tests or metrics do you use to validate a qMR method?

Supporting Information Figure S6. Question 6: What software / packages do you use (or contribute to) for quantitative analysis?

Development, Validation, Qualification and Dissemination of Quantitative MR Methods: Overview and Recommendations by the ISMRM Quantitative MR Study Group

Sebastian Weingärtner, PhD*[,1]; Kimberly L. Desmond, PhD*[2,3]; Nancy Obuchowski, PhD[4]; Bettina Baessler, MD[5]; Yuxin Zhang, PhD[6,7]; Emma Biondetti, PhD[8]; Dan Ma, PhD[9]; Xavier Golay, PhD[10]; Michael A. Boss, PhD[11]; Jeffrey L. Gunter, PhD[12]; Kathryn E. Keenan, PhD[13]; Diego Hernando, PhD[6,7]; on behalf of the ISMRM Quantitative MR Study Group

1. Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands
2. Brain Health Imaging Centre, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
3. Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada
4. Department of Biostatistics, Cleveland Clinic, Cleveland, OH, USA
5. Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Zurich, Switzerland
6. Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA
7. Department of Radiology, University of Wisconsin-Madison, Madison, WI, USA
8. Department of Neurosciences, Imaging and Clinical Sciences, Gabriele d'Annunzio University of Chieti and Pescara, Chieti, Italy
9. Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA
10. Brain Repair & Rehabilitation, Institute of Neurology, University College London, United Kingdom
11. American College of Radiology, Philadelphia, PA, USA
12. Department of Radiology, Mayo Clinic, Rochester, MN, USA
13. National Institute of Standards and Technology, Boulder, CO, USA

Word count: 7500

Corresponding author:

Diego Hernando, PhD
Departments of Radiology and Medical Physics
University of Wisconsin-Madison, Madison, 53705, USA
Email: dhernando@wisc.edu
Twitter: @dherarr

* Sebastian Weingärtner and Kimberly L. Desmond contributed equally to this work

Abstract:

On behalf of the International Society for Magnetic Resonance in Medicine (ISMRM) Quantitative MR Study Group, this article provides an overview of considerations for the development, validation, qualification, and dissemination of quantitative MR (qMR) methods. This process is framed in terms of two central technical performance properties, i.e., bias and precision. Although qMR is confounded by undesired effects, methods with low bias and high precision can be iteratively developed and validated. For illustration, two distinct qMR methods are discussed throughout the manuscript: quantification of liver proton-density fat fraction, and cardiac T_1. These examples demonstrate the expansion of qMR methods from research centers toward widespread clinical dissemination. The overall goal of this article is to provide trainees, researchers and clinicians with essential guidelines for the development and validation of qMR methods, as well as an understanding of necessary steps and potential pitfalls for the dissemination of quantitative MR in research and in the clinic.

Keywords: Quantitative, bias, precision, confounding factors, PDFF, T_1

INTRODUCTION

Magnetic resonance probes a wide array of tissue contrasts, spectral properties and anatomical information. Based on this wealth of contrast mechanisms, a variety of quantitative MR (qMR) methods that extract quantifiable information from MR acquisitions (1-3) have been proposed and continue to emerge from the MR research community. Upon successful development and validation, qMR methods enable improved standardization in the detection, staging, and treatment monitoring of diseases, both in research and in clinical practice (4-7). On behalf of the International Society for Magnetic Resonance in Medicine (ISMRM) Quantitative MR Study Group, we provide an overview of the process of development of qMR methods, as well as guidelines for their technical validation, clinical qualification, application, and dissemination. To illustrate this process, we provide examples from two distinct qMR methods: quantification of liver proton-density fat fraction (PDFF), and T_1 quantification in the myocardium (cardiac T_1 mapping; see Figure 1). These two methods were selected based on their substantial interest within the MR research community, important existing and potential applications, and major advances toward widespread clinical use. Importantly, the current status of development and remaining challenges are different for these two methods, which helps illustrate the diversity in the field of qMR.

Liver Proton-Density Fat Fraction Quantification: Proton-density fat fraction (PDFF) has been developed, validated, and applied for the assessment of tissue triglyceride concentration (8). Although chemical shift encoded (CSE) fat-water imaging was introduced nearly 40 years ago (9), the development of quantitative techniques that measure PDFF has accelerated over the past two decades (10-17). Using either MRS (16) or MRI (8) acquisitions, PDFF measures the concentration of MR-visible triglyceride protons relative to all MR-visible protons (from triglycerides and water), which has multiple research and clinical applications. Recent technical developments and validation studies (see below) have led to widely available techniques. These techniques are particularly promising for the quantification of liver fat, e.g., in the assessment of non-alcoholic fatty liver disease (NAFLD).

Cardiac T_1 Mapping: Although cardiac T_1 mapping was first developed in the 90s (18), the field accelerated more recently when the promise to enable non-invasive assessment of diffuse fibrosis emerged (19). As the T_1 relaxation time depends on the mobility in the macromolecular

environment, over time cardiac T_1 mapping has proved useful in many clinical applications (20). Initially, semi-quantitative relaxation measurements in the myocardium based on Look-Locker sequences were explored (21,22). However, these methods lacked the reproducibility and reliability to serve as a quantitative tool in clinical application. With the introduction of the Modified Look-Locker Inversion Recovery (MOLLI) (23) and shortened MOLLI (shMOLLI) (24) methods, myocardial T_1 mapping became feasible on a voxel-by-voxel basis in a single breath-hold and with high visual T_1 map quality. This facilitated the widespread use and application to numerous ischemic and non-ischemic cardiomyopathies (25,26). Continuous method development and refinement have led to increasingly sensitive and reliable T_1 measurements of the heart, paving the way for routine clinical use (20).

TECHNICAL PERFORMANCE OF QMR METHODS

The development and validation of qMR methods requires a framework for describing their technical performance. The two major technical performance properties of a qMR method (Figure 2) are: i) the bias, which includes the properties of linearity, the regression slope and intercept, and the fixed bias, and ii) the precision, which is described by the repeatability and reproducibility. These metrics are described in detail below and summarized in Table 1. Previous works in qMR have used deviating terminologies, including "accuracy" or "robustness," to describe technical performance. However, in this work we use, and encourage others to use, metrics based on bias and precision, as established by the quantitative imaging metrology community (27,28). A glossary of the terminology used throughout this paper can be found in Supporting Information Table S1.

Bias

Bias describes the systematic tendency of qMR measurements to differ from the ground-truth value of the measurand (i.e., the underlying quantity of interest). To define bias, let X_i denote the ground-truth value of the measurand for the i-th subject, and Y_{ij} denote the j-th qMR measurement for the i-th subject. Ideally, the measurements and the ground truth possess an affine linear relationship, as follows (29,30):

$$Y_{ij} = \beta_0 + \beta_1 X_i + \epsilon_{ij} \quad (1)$$

where β_0 is the intercept, β_1 is the regression slope, and ϵ_{ij} is a random effect, which we assume is independently and identically distributed from a normal distribution with mean zero and variance σ^2 (which captures the precision).

To measure bias, the ground-truth value can sometimes be ascertained by using a reference method. Although the ground-truth should be conceptually well defined, its estimation via a reference method is generally imperfect and requires careful design. Such a reference method may be invasive or non-invasive, and based on MR or other modalities. Importantly, a reference method should be independent from the qMR method under evaluation (e.g., should not be obtained from the same source data), as any dependence between the two measurements may lead to an underestimation of the qMR method's bias. Furthermore, to be accepted as a reference method, its measurements must be highly concordant with the ground truth, and its performance (bias and precision) must be substantially better than the performance of the method under evaluation (28). These requirements often complicate the acquisition of a reference method in vivo.

For this reason, investigators often rely on reference objects ("phantoms") to assess bias. The design of phantoms is driven by the technique they will be testing and the specific tissues or MR properties they will mimic, as well as additional considerations such as traceability and long-term stability (31). It is important that phantoms themselves are systematically measured prior to use, which is sometimes achieved using gold standard NMR measurements (32), the best available reference method on MRI systems, or non-MR methods. With a standard phantom, such as the ISMRM/NIST system phantom (33), or a well-characterized home-built phantom, the technical performance of a qMR method can be estimated, as an approximation of in vivo technical performance. Although phantoms are highly effective in many qMR applications, there are cases where phantoms may be of limited value, as existing phantom designs do not adequately replicate the relevant signal properties, spatial distribution, or temporal dynamics found in tissue (34).

In a phantom study (or in vivo, if a suitable reference method is available), measurements are obtained at multiple values X_i (e.g., corresponding to different phantom compartments) over the range of the true value, X. Ideally, at least 10 nearly equally spaced values X_i should be chosen (29), covering the range of values of interest. This range usually includes the normal range expected in a healthy reference cohort as well as values observed under influence of the pathology or condition of interest. For each value i, the individual bias or % bias is calculated as:

$$b_i = (\overline{Y}_i - X_i)\,; \qquad \%b_i = \left[(\overline{Y_i} - X_i)/X_i\right] \times 100 \qquad (2)$$

where $\overline{Y_i}$ is the mean over the potentially repeated measurements on the same phantom compartment (or subject). Finally, some qMR methods may present a constant bias that is not dependent on the true value of the measurand. In these cases, the fitted line in Eq. 1 will be parallel to the identity line, with regression slope β_1 close to one, and regression intercept β_0 that provides an estimate of the bias (29).

Over N observations, we can estimate the fixed bias:

$$\hat{b} = \sum_{i=1}^{N} \%b_i / N \qquad (3)$$

A 95% confidence interval (CI) for the fixed bias should be reported with the estimate of bias (35). Note that small values of fixed bias are often well tolerated. For example, under typical conditions, confidence intervals for a new patient's measurement constructed under the no-bias assumption provide nominal coverage as long as the fixed bias is <12% of the wSD (36).

Because the bias sometimes depends on the true value of the measurand, the bias profile can be evaluated by plotting the estimate of bias from each value of X_i (i.e., b_i) against the true values X_i. Note that the relationship between measurements and ground truth may generally be nonlinear, particularly when considering a broad range of measurand values. However, the assumption of linearity is often an appropriate approximation and greatly simplifies the statistical analysis. To assess the property of linearity, we fit an ordinary least squares (OLS) regression of the Y_{ij}'s on X_i's. One way to test the appropriateness of a linear model is to formally test for significant non-linearities (curvature) (29). Sequential tests can be performed starting with a third-order (cubic) regression: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$. If the third order coefficient β_3 is not significantly different from zero, then the process can be repeated with a second-order (quadratic) regression: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$. If the quadratic term β_2 is not significantly different from zero, then the hypothesis of a linear model cannot be rejected, and a linear fit can be used: $Y = \beta_0 + \beta_1 X$ (29). Ideally, R-squared (R^2) will be greater than 0.90 (35). The regression slope, β_1, should be reported along with its 95% CI. As a general rule, a slope in the range [0.95, 1.05] is acceptable (35). Sometimes, the linear relationship in Equation 1 holds only for a certain range of the values of the measurand, so it is important to assess this property over the likely values of the measurand. It may be that linear relationships hold for various ranges of the true value, but β_0 and β_1 differ for each range or even vary by subject characteristics (e.g., age or body mass index).

Finally, once the bias is known, the qMR method could, in principle, be calibrated to the reference to eliminate bias. However, this is not a common approach. Indeed, the bias itself often arises from uncorrected confounding factors that may affect various acquisitions or patients differently (37,38). For this reason, bias is often not reproducible, and calibration-based correction should be approached with caution.

Precision

Precision describes the tendency of the measurement system, when used repeatedly in several "replicate" measurements on the same subject, to produce different values (27). The precision of a method has enormous practical importance. Indeed, the required number of participants for clinical studies increases with $\sigma^2$and is therefore driven by the precision of a qMR method, which determines their cost and feasibility (39). The precision is also a major factor affecting the method's ability (including sensitivity and specificity) to detect a specific condition, and determines the minimum detectable change (see Figure 3). In contrast to bias, the evaluation of precision does not require a reference method, and therefore in vivo evaluation of precision is often highly feasible.

Some studies use spatial variability in a homogeneous phantom or tissue as a heuristic to evaluate precision metrics. While this may be an acceptable approximation with certain simple imaging methods, spatial variability of system properties (e.g. B_0 and B_1^+ heterogeneities) often render this approximation inadequate, even if the region of interest appears homogeneous to the observer. In particular, this spatial variability method cannot be used to evaluate precision metrics if spatial information is used in the image reconstruction or parametric mapping (e.g., regularization in compressed sensing). Instead, precision metrics should be evaluated in studies that obtain and compare multiple replicate measurements.

Test-retest studies allow estimation of precision. When the same MR system and experimental conditions (including acquisition parameters) are used for all replicate measurements on the same subject over a short span of time, we refer to this as the *repeatability* condition. When the replicates are obtained under different conditions (e.g., different field strengths, different MRI vendors, platforms, or software versions, different individual scanners, different pulse sequences or acquisition parameters, different image analysis software, different readers, or long delay between acquisitions), we call this the *reproducibility* condition [3]. With qMR, we often characterize

precision by either the *within-subject standard deviation*, denoted wSD, or the *within-subject coefficient of variation*, denoted wCV. Note from Equation 1, $wSD = \sqrt{\sigma^2}$, and $wCV = \sqrt{\sigma^2}/\overline{Y}_i$. The wCV is often used when the variability in the measurements is much higher for large true values or when the measurements are log-normally distributed (thus, X_i and Y_{ij} in Equation 1 would need to be measured on a logarithmic scale).

Precision studies are often small because of cost, ethical, and technical concerns (29,30,40). For these reasons, meta-analysis is often required to pool estimates from multiple studies (41). A general rule of thumb to obtain a reliable estimate of precision is >35 subjects with two or more replicates (36).

For each subject in a test-retest study, the qMR measurement is performed at time point 1 (denoted Y_{i1}) and time point 2 (Y_{i2}). Additional time points can be included, if available. For each subject, we can calculate the mean and SD of the measurements: $\overline{Y}_i = (Y_{i1} + Y_{i2})/2$ and $SD_i = \sqrt{(Y_{i1} - Y_{i2})^2/2}$. From the N subjects, we estimate the mean wSD or wCV as:

$$\widehat{wSD} = \sqrt{\sum_{i=1}^{N} SD_i^2/N}; \qquad \widehat{wCV} = \sqrt{\sum_{i=1}^{N} \left(SD_i^2/\overline{Y}_i^2\right)/N} \qquad (4)$$

Importantly, 95% CIs for wSD and wCV should also be reported (35). Implicit in Equation 4 is the assumption that wSD (or wCV) is constant over the range of measurand values. This assumption should be assessed by calculating the estimates for several ranges of $\overline{Y}_i$ or even for various patient and/or disease characteristics, to determine a precision profile (42,43).

Two useful precision metrics are the repeatability coefficient (*RC* and *%RC*), estimated as:

$$\widehat{RC} = 2.77 \times \widehat{wSD}; \qquad \widehat{\%RC} = 2.77 \times \widehat{wCV} \times 100 \qquad (5)$$

and the reproducibility coefficient (RDC and %RDC), estimated analogously. These metrics describe the smallest significant difference between two repeated measurements on a subject and assuming a normal distribution for the replicate measurements (27,30). These metrics can be used as thresholds to discern between differences due to measurement imprecision and differences due to a true change in the measurand.

In evaluating reproducibility, the experimental conditions across replicate measurements can be altered in various different ways (see above). Ultimately, the widespread dissemination of a

qMR method will require establishment of reproducibility across conditions such as different centers, MRI vendors, or patient populations. However, such multi-center, multi-vendor studies are expensive and complex, and may not be appropriate for a newly developed qMR method. A practical approach is to evaluate reproducibility in multiple studies of increasing complexity, beginning with relatively simple studies at a single center and vendor (44), while building up toward more ambitious studies (45).

In evaluating precision, it is important to carefully design and describe the specific procedures followed. For example, repeatability may be evaluated by performing consecutive scans within the same scanning session, in order to establish the effects due to MR system adjustments and noise. However, repeatability is often determined by scanning the subject in separate sessions over a short time interval, including repositioning and re-localizing between sessions, in order to capture additional variability due to other factors such as subject positioning (29). In general, the experimental design to study precision may be different for different methods or applications. Thus, it is important to meticulously report the parameters and conditions that are kept identical and those which may have changed between replicate measurements, to enable replication and interpretation of the results.

Examples

Liver PDFF Quantification

Liver PDFF quantification methods have been validated in multiple studies including evaluation of bias in PDFF phantoms (both commercially available and home-built) (46,47), in vivo liver imaging (48), and ex vivo livers (49). In a recent meta-analysis (48), liver PDFF had high linearity and low bias with respect to the MRS-determined reference PDFF value in 23 studies, which included a total of 1,679 subjects. Test-retest repeatability studies have also been performed (48,50). Recently, high linearity and low bias of multi-center PDFF measurements (47) has been demonstrated by shipping a phantom to multiple centers in a “round-robin” study, and evaluating measurements on the same phantom across centers, vendors, platforms, field strengths, and acquisition parameters. In addition, the reproducibility of PDFF measurements in the liver, including across field strengths and MRI vendors, has also been demonstrated in multiple studies (48,51).

Cardiac T_1

Bias and precision have been the dominating criteria in analysis of cardiac T_1 mapping methods [35]. Multiple studies have shown that different T_1 mapping methods provide varying profiles of bias and precision.

Inversion recovery-based methods have been shown to exhibit good repeatability but large biases, while saturation recovery methods have been shown to reduce bias at the cost of reduced repeatability (52-54). For example, the most commonly used myocardial T_1 mapping technique, the inversion recovery-based method MOLLI, is known to be subject to multiple confounding factors and exhibits substantial bias (55). However, given its excellent repeatability and visual image quality, the sequence is highly popular among users (23). Even though some studies have shown initial evidence of multi-center or multi-vendor reproducibility with tightly controlled protocols (56), the reproducibility is generally compromised due to the measurand confounders (see next section). Thus, it is recommended to obtain center and protocol specific reference ranges in healthy subjects, before using MOLLI for quantitative diagnosis (20).

Table 1: Technical performance metrics.

Metric	Definition	Liver PDFF	Cardiac T_1 (MOLLI)
Linearity	Ability to provide measurements that are proportional to the true value as described in Equation 1	r^2 =0.96 and no evidence of significant higher-order terms in regression analysis of measurements vs true value (48).	r^2 = 0.996, magnitude of higher order terms < 0.0001(52).
Regression slope	β_1 in Equation 1	0.975 (48).	0.919 (52).
Fixed bias	Difference between average measurement and true value	<0.2% (48).	4.2% (52).
Bias (or precision) profile	A table or figure illustrating the estimates of the bias (or precision) over the range of true values and/or other relevant characteristics	See Yokoo et al., 2018 (48).	See Roujol et al., 2014 (52,57).
Repeatability	A measure of precision describing the variability in measurements on a subject over a short period of time using the same imaging system and experimental conditions (27)	Repeatability coefficient of 2.9% (48).	Repeatability coefficient of 2.0% (52) - 4.6% (58)
Reproducibility	A measure of precision describing the variability in measurements on a subject using different experimental conditions (different systems, and/or pulse sequence parameters, and/or measurements separated by a long period of time, etc) (27)	Reproducibility coefficient (across different hardware systems or reconstruction software) of 4.3% (48).	Highly variable. In tightly controlled studies, 2.1% has been reported (59), however, a meta-analysis showed >7% reproducibility in healthy subjects (60). For this reason, it is not recommended to compare MOLLI T_1 values across systems and parameters, due to system specific biases (20).

CONFOUNDING FACTORS IN MR

In qMR, a wide variety of confounding factors may introduce bias or poor precision. Table 2 provides illustrative categories and examples of qMR confounding factors. A poll distributed among the members of the ISMRM Quantitative MR Study Group queried the frequency, relevance, and potential correction mechanisms for confounding factors used in the quantitative MR community. Supporting Information Figures S1-S6 summarize the poll results.

Table 2: Types of qMR confounding factors and illustrative examples

Hardware/system imperfections	**Physiological effects/motion**	**Signal model imperfections**	**Other artifacts and noise**
B_0 heterogeneities and off-resonance	Respiratory motion	Additional relaxation mechanisms (not included in model)	Partial volume
B_1 heterogeneities	Cardiovascular motion/pulsation		Slice profile imperfections
Eddy currents	Intestinal peristalsis	Spectral complexity (additional resonances, J-coupling, etc.)	Imperfect spoiling
Gradient nonlinearities	Bulk body motion		Parallel imaging artifacts
System drift	Blood flow	Exchange (multi-pool)	Noise

Hardware and system imperfections

The presence of magnetic field heterogeneities (both B_0 and B_1) (61), gradient nonlinearities (62), concomitant gradients (63,64), eddy currents (65), system drifts (66), timing errors (67), and other system imperfections, is unavoidable in MR applications. These effects may result in tolerable artifacts in qualitative MR as long as the relative visual contrast between tissues is preserved, but may introduce substantial bias and poor precision in qMR methods.

Physiological effects and motion

Physiological motion effects include respiration, cardiovascular motion and pulsation, intestinal peristalsis, and bulk patient motion, among others. These effects often result in artifacts, ghosting,

and mis-registration in the acquired images (68), which can in turn introduce bias and poor precision in qMR. Blood and tissue motion during the acquisition can also introduce artifacts, phase offsets, and dephasing that confound the quantification.

Signal model imperfections

Practical qMR methods rely on simplified signal models. The presence of signal effects that are not included in the model introduces bias in qMR measurements. These effects may be due to additional relaxation mechanisms, incomplete approach to steady state, diffusion, spectral complexity, etc. Signal model imperfections can lead to poor reproducibility in qMR, as these effects will often manifest differently for varying experimental conditions, including different systems, field strengths, and acquisition parameters. Ideally, signal models should be based on specific biophysical assumptions about the tissues of interest. However, biophysical modeling is challenging in certain applications, and so signal "representations" are often used, which enable fitting of the acquired data but are not based on specific tissue models (69). For example, the diffusion tensor representation provides a useful approximation to the diffusion-weighted MR signal at moderate *b*-values, but is not based on specific tissue modeling assumptions (69). Such signal representations have demonstrated clinical value, but their quantitative performance needs cautious consideration. For example, bias may not be meaningful if the measurand does not represent a physical property of the tissue, and reproducibility across changes in acquisition parameters is often challenging.

Other artifacts and noise

A variety of additional imaging artifacts, including partial volume, slice profile imperfections, imperfect spoiling, parallel imaging artifacts, and noise can confound qMR methods. For example, imperfect slice profiles due to finite-duration excitation pulses lead to a distribution of flip angles across the slice and may also introduce crosstalk between slices (70,71). In addition, noise in the acquired imaging data propagates into the subsequent qMR measurements. The propagation of noise is generally dependent on the acquisition parameters and the choice of signal model; more complicated models with many free parameters often result in higher noise amplification. Further, manipulation of MR signals prior to qMR measurement can affect the noise distribution, which affects the bias and precision of qMR methods. For example, noise in complex MR data is well

modeled by a Gaussian distribution. However, qMR sometimes relies on magnitude images (e.g., in diffusion MRI, or cardiac T_1 mapping as discussed throughout this paper). This magnitude operation has several important effects, including the elimination of phase information, and the introduction of an additional bias. Indeed, regions of low signal magnitude, as commonly observed in methods such as diffusion MRI or relaxometry, deviate substantially from a Gaussian noise distribution (72). If subsequent qMR processing implicitly assumes a Gaussian noise distribution (e.g., in methods that rely on least-squares fitting, as described below), bias and poor reproducibility may result from the inaccurate noise assumptions (73,74).

For these reasons, different processing pipelines of the same data can lead to differences in the resulting qMR measurements. For example, even filtering of the image data can introduce biases in the quantification when nonlinear models are being used. Thus, using transparent open-source toolboxes or custom-built processing for qMR should be preferred over black-box tools, when reproducibility is targeted.

Finally, when using MR-based reference methods for validation of qMR, even the reference method itself may not be immune to the presence of confounding factors such as physiological effects and motion. This limitation of the reference method may complicate the evaluation of qMR bias in vivo.

Examples

Liver PDFF Quantification

Quantification of PDFF is affected by multiple confounding factors, including:

- **T_1 recovery:** The short T_1 relaxation time of fat compared to that of water in the liver can lead to bias (overestimation) of PDFF (12) in acquisitions that include T_1 weighting.
- **T_2* relaxation:** T_2* decay across multiple echoes can appear as interference between fat and water signals, and therefore can introduce bias and poor reproducibility in PDFF quantification (see Figure 4) (13,14,75,76).
- **Spectral complexity of fat signals:** Unlike water signals, which result in a single MR resonance, fat signals arise from protons located in various positions within the triglyceride molecule. These protons, in turn, lead to a multi-peak spectrum from fat (15,77). If unaccounted for, this spectral complexity leads to bias and poor reproducibility across acquisition parameters, particularly the echo time combination.

Phase errors: Phase errors, such as those arising from eddy current effects, can introduce bias and poor reproducibility in PDFF quantification (78,79).

Over the past two decades, these and other confounding factors have been systematically identified, characterized and addressed using various acquisition and post-processing-based approaches (see the Technical Development section below).

Cardiac T_1

Quantitative cardiac imaging is particularly challenging due to the impact of cardiac and respiratory motion. Confounders, as described above, often have an imaging method-specific impact on the quantification. Additionally, subject-specific effects are often substantial. Various existing and emerging technical developments seek to mitigate these effects (55,57,80-83). The most relevant confounders for cardiac T_1 mapping include:

Heart rate: Acquisitions commonly need to be synchronized with the heartbeat to minimize cardiac motion effects. The subject-specific heart rate may influence the bias and/or precision. Introduction of alternative mapping schemes or heart rate-resilient timing has helped to alleviate this confounder (55,57,80).

k-Space acquisition: In some methods, only the effect of the magnetization preparation is modeled. In this case the disruption of the magnetization by RF pulses used for the k-space acquisition can cause bias in the quantification. Specifically, this may render the quantification susceptible to physiological factors such as T_2 relaxation (57) or magnetization transfer (84), or system-related properties such as off-resonance (85) or flip angles (86). The dependency on system-related parameters makes it paramount to use identical sequences and sequence parameters in cardiac T_1 mapping, when reproducibility across centers and MRI scanners is desired.

Partial-volume effect: Cardiac acquisitions are commonly limited in the achievable resolution due to motion constraints. This results in partial-voluming, where voxels are partially filled with different tissue types at tissue interfaces. As a result, the area that can reliably be evaluated is further reduced, rendering the quantification dependent on accurate delineation of the region of interest (87,88).

TECHNICAL DEVELOPMENT AND VALIDATION

The development of qMR methods is typically an iterative process including design of acquisition, modeling and signal fitting methods. This technical development can be framed as an optimization of bias and precision metrics (see Figure 5), subject to specific constraints such as scan time, or hardware performance.

Acquisition

Once the basic physical mechanism to be probed has been selected and potential confounders have been identified, aspects of protocol design and optimization can be considered. A common goal is to select pulse sequences and parameters such that the measurand can be determined with low bias and high precision, subject to a set of timing, hardware, and other constraints. Acquisition design will often begin by selecting a pulse sequence where the measurand of interest can be directly probed, while minimizing the effect of confounding factors. Next, an acquisition that includes multiple scans with different parameters can be designed to enable estimation of the measurand. In applications where thermal noise is the dominant source of noise (as opposed to, e.g., physiological noise), acquisitions with higher imaging SNR may substantially improve bias (12) or precision (89). The choice of acquisition parameters may be driven by heuristics, and also refined using quantitative tools such as sensitivity analysis (90,91), or noise propagation analysis (e.g., Cramer-Rao lower bounds, CRLB) (11,92-94).

Liver PDFF Quantification

The choice of pulse sequence for quantification of PDFF is driven by the desire to obtain chemical shift-encoded data with proton-density contrast (e.g., avoiding confounding effects due to T_1 and T_2 relaxation), and with rapid scan times (e.g., to enable whole-liver coverage in a single breath-hold while avoiding motion artifacts). For these reasons, the pulse sequence of choice for MRI-based liver PDFF quantification is a multi-echo spoiled gradient echo (SGRE) sequence, either using 2D multi-slice or 3D imaging (8,17). In addition, small flip angles are used in order to avoid T_1 bias (12). Other confounding factors are typically addressed by postprocessing/modeling (see below). Optimal acquisition parameters, such as echo times, have been determined using CRLB analysis (11).

Cardiac T_1

A multitude of pulse sequences for cardiac T_1 mapping have been proposed and novel method development remains an active area of research. Generally, these acquisitions can be decomposed into three integral parts: 1) contrast sensitization; 2) k-space acquisition; 3) motion compensation.

Typically, preparation pulses are used to sensitize the imaging signal to the T_1 relaxation time of the tissue. Inversion pulses are most widely used in T_1 mapping, including the commonly used MOLLI sequence and its variants (23,24,55). Saturation recovery has also been proposed with the potential to minimize bias caused by various confounding factors as described above (57,80). However, due to a decreased dynamic range, saturation recovery preparation typically results in lower T_1 mapping precision as compared with inversion preparation.

Cardiac T_1 mapping is typically performed using multiple electrocardiogram (ECG)-triggered snapshot images, with each image obtained during a single diastolic quiescence, i.e., all *k*-space lines necessary for image reconstruction of one snapshot image are acquired in one heartbeat. To achieve optimal SNR as well as minimal disruption of the longitudinal magnetization recovery curve, balanced Steady State Free Precession (bSSFP) readouts are the method of choice. Spoiled gradient echo readouts have also been explored to minimize sensitivity to off-resonance and field heterogeneities, albeit at the cost of reduced precision (95). More recently, continuous imaging throughout the heartbeat have been proposed to allow cardiac phase-resolved T_1 mapping (81,82,96).

ECG triggering is almost universally used as the means for cardiac motion compensation in T_1 mapping, with few notable exceptions (96,97). Various schemes have been explored for respiratory motion compensation. Clinically available T_1 mapping methods usually acquire a single-slice T_1 map in a single breath-hold (26). However, free-breathing methods have also been explored with diaphragmatic navigator gating, tracking or self-gating (97-99). Importantly, free-breathing sequences allow for the acquisition of multiple slices or 3D volumes and can be used to enable T_1 mapping with increased spatial resolution (83,100,101).

Model selection

Many qMR methods rely on parametric mapping using a signal model that relates the acquired data to the underlying measurand. Selection of a signal model is typically an iterative process and seeks to balance bias and precision. Often the process begins with identifying the relevant degrees

of freedom in the underlying tissue (69), such that these tissue properties can be related to, and estimated from, acquired MR signals. For example, this step may involve the identification of the major pools of nuclei with shared properties that will reasonably contribute to the signal. These pools can describe physical compartments, such as "intracellular" compartments, or local molecular environments such as lipid protons. The Bloch equations, describing the response to RF energy and relaxation properties, are defined for each pool. Next, a model may consider whether nuclei can travel between pools by chemical exchange or diffusion, or interact magnetically with other pools due to proximity and thus define the exchange kinetics. Models commonly describe the signal within a voxel independently of its spatial neighborhood, but one may also need to consider the influence of the neighboring voxels (e.g., in quantitative susceptibility mapping (102), or electrical properties tomography (103,104)).

The next step is often to evaluate the signal model under modifications of the acquisition pulse sequence. It is often helpful to develop a working model for simple excitation-readout with Cartesian acquisition and long TR before considering advanced *k*-space trajectories or pulse trains. The requirements for each measurand are different, but major considerations in the presence of increasingly advanced pulse sequences may include relaxation effects, B_0 and B_1 heterogeneities, etc. It may also be necessary to consider the need for steady-state or non-steady state modeling. Finally, any signal manipulations that occur before analysis, such as magnitude operation or spatial filtering, need to be included.

In subsequent iterations, confounding factors are often identified and addressed through acquisition- and/or modeling-based refinements. Importantly, qMR methods necessarily use simplified models of the actual underlying physics. For this reason, it is always possible to "enhance" the models by including additional unknown parameters. However, these signal model enhancements generally lead to increased challenges in the parameter estimation (particularly noise amplification, sensitivity to artifacts, and computation time). Practical signal models, therefore, seek a balance between accurately capturing the underlying physics and enabling stable quantification within acceptable computation times. Once a satisfactory model is achieved, this model often needs to be re-evaluated upon subsequent refinements of the qMR method, including accelerated acquisitions.

Liver PDFF Quantification

A widely used signal model for PDFF quantification in the liver includes (8,15,105):

$$s(TE_n) = \left(\rho_w + \rho_F \sum_{m=1}^{M} \alpha_m e^{i2\pi f_{F,m} TE_n}\right) e^{i(\phi_0 + 2\pi f_B TE_n)} e^{-TE_n/T_2^*} \quad (6)$$

where ρ_w and ρ_F are the proton density-weighted signal amplitudes of water and fat, respectively; fat signals are modeled as a pre-calibrated spectrum including M peaks with known relative amplitudes α_m and frequency offsets $f_{F,m}$ (77); initial phase ϕ_0; B_0 related off-resonance frequency f_B; and transverse relaxation time T_2*. Upon data fitting (see below), this signal model allows estimation of ρ_w and ρ_F, which lead to the calculation of PDFF as:

$$PDFF = \frac{\rho_F}{\rho_w + \rho_F} \quad (7)$$

Importantly, the widely used signal model in Equation 6 constitutes a balance between bias and precision (noise performance) (106). For example, this model addresses the spectral complexity of the fat signal by using a multi-peak signal model and also accounts for T_2* decay. If unaccounted for, both of these effects have been shown to lead to substantial bias and poor reproducibility in PDFF quantification (17). However, the model in Equation 6 typically relies on a pre-calibrated multi-peak fat spectrum, where the relative frequencies and amplitudes of the fat peaks are assumed known a priori (107), and also assumes a common T_2* decay time for water and all fat peaks (106). Each of these approximations help maintain acceptable noise performance and precision for PDFF quantification by limiting the number of unknown parameters, even though they may introduce a small bias in the estimation of liver PDFF when the model assumptions do not hold exactly.

Cardiac T_1

T_1 recovery is thoroughly studied and can be accurately described by the well-known phenomenological Bloch relaxation equations. However, the signal model needs to be adapted to the specific imaging sequence, as summarized next.

In the widely used Modified Look-Locker Inversion Recovery (MOLLI), an adaptation of the standard inversion recovery model is commonly employed to describe the signal $S(t)$ at different inversion times t:

$$S(t) = A - Be^{-t/T_1^*}. \quad (8)$$

Here A and B describe fit parameters and T_1^* is the apparent relaxation time. The T_1 estimate is then extracted as $T_1 = \left(\frac{B}{A} - 1\right)T_1^*$. This adaptation is inspired by Deichmann et al. (108) and aims to reduce the effect of the RF pulses used for the k-space acquisition on the quantification. However, the acquisition commonly deviates from the assumptions underlying this correction, inducing residual susceptibility to various effects related to the RF pulses during the k-space acquisition. Numerical models have also been proposed to approximate the magnetization evolution during the k-space acquisition, using for example Bloch equation simulations or additional parameters (109,110). While these methods commonly achieve lower bias, their applicability might be limited, and precision may be compromised.

A second class of myocardial T_1 mapping methods is based on saturation recovery, which commonly relies on the following three-parameter model:

$$S(t) = A\left(1 - e^{-t/T_1}\right) + B \qquad (9)$$

This saturation recovery-based approach has been shown to compensate for the effects of the RF pulses used for the k-space acquisition (57), which in turn enables cardiac T_1 mapping with reduced bias. It has also been suggested to omit the B parameter to obtain a two-parameter model, in order to improve precision at the cost of bias in saturation recovery T_1 mapping (55).

Model Fitting

Fitting a signal model to acquired data can typically be described in terms of a formulation and an algorithm, as described next.

The formulation is often an optimization problem, which describes in what sense the model should fit the acquired data. For example, least-squares fitting is often used for various linear or nonlinear models in quantitative MR. For nonlinear models based on exponential functions, a logarithm of the data is sometimes calculated to linearize the problem. This linearization simplifies the optimization, although it affects the noise propagation and may require additional manipulations to avoid excessive noise influence from low-SNR data points (111). Further, some formulations rely on the acquired complex data, whereas others use magnitude data (79). Finally, the formulation may be constrained (where the set of allowable parameters is restricted based on physical or noise propagation considerations) or unconstrained. In addition to least-squares fitting,

other formulations can be used, including those required for maximum-likelihood estimation in the presence of non-Gaussian noise.

Once a formulation is selected, an algorithm needs to be selected to solve the corresponding optimization problem. Depending on the formulation, various closed-form or iterative algorithms are typically available. Variations of Newton's method, including Levenberg-Marquardt and Gauss-Newton algorithms, constitute common choices for iterative optimization (112). An ideal algorithm would be efficient (i.e., fast and requiring low resources) and would lead to the global solution of the optimization problem described in the formulation.

Liver PDFF quantification

Model fitting for PDFF mapping is typically performed using nonlinear least-squares fitting of the signal model (Eq. 6) to the acquired multi-echo data, followed by calculation of PDFF at each pixel (Eq. 7). Multi-echo data are often corrupted by phase errors that are inconsistent across echoes (78,79). For this reason, some or all of the phase information is often discarded to avoid PDFF bias, and algorithms often rely partly on fitting the signal magnitude, instead of the original complex-valued signals. Magnitude fitting leads to reduced bias by avoiding phase related PDFF errors at the cost of reduced noise performance and precision (by discarding half of the acquired information, i.e., the phase).

Cardiac T_1

Basic model fitting in cardiac T_1 mapping is also most commonly performed using magnitude-based nonlinear least-squares fitting. When unsigned magnitude images are used in an inversion-recovery model, the signal polarity information is lost. To resolve this issue the images are commonly ordered by the inversion time and the polarity can be restored heuristically, by successively flipping the sign in the ordered sequence and accepting the solution with the lowest fit residual (20).

However, this process might introduce additional noise variability. It has been proposed to incorporate phase information to perform hybrid fitting on a signed magnitude. Here the background phase is extracted from a fully relaxed image, and the phase difference to other T_1 weighted images can be used to restore the signal polarity (113).

CLINICAL QUALIFICATION

For qMR methods, in addition to technical validation, which measures the bias and precision of the quantitative measurements, it is essential to perform clinical qualification (Figure 6) (114). Clinical qualification seeks to establish the relationship between the qMR measurement and biological processes or clinical endpoints, as needed to determine the clinical utility of the method, e.g., whether it enables screening, diagnosis, staging, prognosis, or treatment monitoring for a particular condition and target population (115-119). For example, rather than focusing on technical performance metrics of bias and precision, clinical qualification may focus on metrics such as sensitivity, specificity, negative / positive predictive value, prediction accuracy, or odds ratio (114,120). Upon successful clinical evaluation for a specific application, qMR methods may lead to qualified quantitative biomarkers (see Glossary in Supporting Information Table S1) (6).

This marks an important distinction: while a qMR measurand usually relates to a physical property, this measurand, when being used as a clinical biomarker, indicates pathophysiological alterations or other changes in the physiological state. Often numerous biological and physiological processes affect the underlying physical property. Thus, a single qMR measurand can be qualified as a biomarker for multiple disease entities. In this case, while being sensitive to multiple diseases, the measurand may not be specific to any one physiological alteration.

There are strong connections between technical validation and clinical qualification. For example, a qMR method with poor precision (e.g., poor test-retest repeatability) will likely also have poor sensitivity and specificity for detection of a specific condition. However, these are also important distinctions between both types of evaluation. For example, it is possible to develop a qMR method with excellent technical performance (low bias and high precision) for quantifying a measurand; however, this method may have poor clinical performance for a specific application, e.g., due to underlying biological variability that complicates the relationship between the measurand and the clinical endpoint of interest, such as survival, disease-free survival, or various surrogate endpoints (121-123). Further, a biased method may reduce the desired effect size, as the bias itself may be different for various patient populations. Alternatively, it is possible that a biased (confounded) qMR method provides larger effect sizes for a specific disease entity than an unbiased measurement, e.g., if the confounders themselves are sensitive to the physiological alteration (see Figure 7). However, it is important to note that this enhancement usually comes at the cost of strongly reduced reproducibility as the variability in the bias is difficult to control.

For these reasons, it is essential to conduct both technical validation and clinical qualification of qMR methods. This need further highlights the importance of multi-disciplinary collaboration between technical imaging researchers, translation-focused radiologists, and other clinicians.

Examples

Liver PDFF Quantification

Liver PDFF has been shown to be correlated with histologic steatosis grade. For example, PDFF can classify histologic steatosis (grade 0 vs. 1-3) with sensitivity 0.93 and specificity 0.94 (124). Also, MRI-based liver PDFF quantification is emerging as a useful biomarker to assess longitudinal changes in liver fat within clinical trials (125). Further, a reduction of MRI-PDFF by 30% is associated (odds ratio 6.98) with histologic improvement in NAFLD Activity Score (126).

Liver PDFF values may be predictive of pediatric metabolic syndrome (127). In addition, liver steatosis is associated with cardiovascular diseases (128,129); for instance, liver fat is an independent risk factor (odds ratio 2.1) for high-risk plaque (128) and other cardiovascular risk factors (129). Importantly, now that the required MRI technical development is mature and PDFF mapping methods are widely available, determination of the association between liver PDFF and various clinical outcomes constitutes an active area of research.

Cardiac T_1

Native myocardial T_1 times, in the absence of a contrast agent, have been evaluated against histologically determined fibrosis from myocardial biopsies in vivo and total collagen volume in animal studies (130) and heart transplant patients (131). Variable degrees of correlation ranging from moderate to high have been reported depending on the disease model (131), indicating that T_1 times are not reflective of fibrosis alone but of a number of factors. Extracellular Volume (ECV) calculated from native T_1, post-contrast T_1, and hematocrit, generally showed better correlation to the amount of fibrosis but variability among disease models and studies remains. Accordingly, the clinical context needs to be considered in the interpretation of both native T_1 and ECV, and alteration in either measurand cannot be directly linked to a single specific physiological process (20,132-134).

Nonetheless, cardiac T_1 mapping-related markers have demonstrated high clinical diagnostic and prognostic value in an unexpectedly wide range of disease entities (25,135). For example, in cardiac amyloidosis ECV showed excellent sensitivity and specificity (0.93 and 0.87, respectively) and an odds-ratio of 84.6 (136). In patients with an acute infarct quantitative assessment of normal appearing myocardium in patients with an acute infarct using native T_1 or ECV has proven to be a better predictor for all-cause mortality or major cardiac events than any other cardiac MRI marker (131), and accurate differentiator between reversible and irreversible myocardial damage (96.7% prediction accuracy).

Interestingly, MOLLI T_1 mapping, which is most common in clinical use, is known to exhibit a large bias. However, it has been suggested that certain confounders to MOLLI T_1 measurements may enhance clinical sensitivity (137). As illustrated in Figure 7, this somewhat counter-intuitive phenomenon arises because some confounders (e.g., magnetization transfer (84,137,138)) are sensitive to pathological alterations, leading to inflated effect size in certain disease entities compared to unbiased measurements. However, as a result of these confounders, the measurand in MOLLI is highly dependent on the sequence parameters (e.g., TR, flip-angle, slice-profile), the scanner specifications and tissue properties that are not related to the tissue T_1 time. Hence, this inflated effect size is obtained at the cost of reduced reproducibility.

DISSEMINATION IN RESEARCH AND IN THE CLINIC

Although qMR methods have shown great potential to guide clinical decision-making and patient management for improved patient care and outcomes, very few qMR methods are used in routine clinical practice. Many promising qMR methods are only described in the scientific literature, without translation in clinical research studies or clinical use.

As recently stated in the *Imaging Biomarker Roadmap for Cancer Studies* (139), all imaging biomarkers, including quantitative MRI biomarkers, need to cross two "translational gaps" before they are ready to guide clinical decision making. These gaps are crossed through increasing technical validation and clinical qualification, as well as assessment of cost effectiveness and other considerations. Once technical and clinical performance evaluation demonstrate the reliability of a qMR biomarker to test medical research hypotheses, this biomarker can cross the first gap and become a useful "medical research tool". At this stage, substantial additional validation and

qualification are still needed in order to achieve clinical impact as a screening, diagnostic, or predictive biomarker. This may include validation of multi-center reproducibility (140), and large prospective clinical trials to demonstrate improved clinical diagnosis or outcomes. In combination with cost effectiveness and other considerations, these activities enable a biomarker to cross the second translational gap and become a "clinical decision-making tool" that influences patient care. Importantly, application of qMR methods in research or in the clinic requires rigorous quality assurance and quality control procedures (141-143).

During the three processes of i) technical development and validation, ii) clinical qualification, and iii) dissemination, additional substantial challenges (including regulatory issues and market-related factors) often arise before clinical dissemination is achieved (Figure 6). For example, the ability of healthcare providers to obtain reimbursement or take charge of the costs associated with quantitative imaging biomarkers may drive the clinical use of these tools. Oftentimes, a lack of CE/FDA (or equivalent) labelling limits the ability to apply a biomarker in clinical practice.

Examples

Liver PDFF Quantification

CSE-based liver PDFF quantification has emerged as a major clinical and research tool to determine liver fat content. Importantly, this qMR method is commercially available on systems from various MRI vendors, including regulatory approvals such as FDA clearance and CE mark. A liver PDFF quantification profile is currently being developed by the Radiological Society of North America's Quantitative Imaging Biomarkers Alliance (RSNA QIBA) (144). Now that liver PDFF quantification methods have shown excellent technical performance (low bias, high precision), clinically relevant results are emerging, including population studies measuring prevalence in various populations (145) and clinical studies showing the prognostic value of PDFF (146).

Cardiac T_1

In recent years, cardiac T_1 mapping has become widely available on most clinical MRI systems. Some vendors have released dedicated product packages comprising one or more T_1 mapping methods, while others have provided prototype methods. Several cardiac T_1 mapping methods have regulatory approval such as FDA clearance and CE mark.

T_1 mapping is widely used in cardiac MRI in academic centers and beyond. It has been successfully applied to an unexpectedly large spectrum of ischemic and non-ischemic cardiomyopathies (20,131) and is established as part of routine scanning in numerous clinical cardiac MR protocols. The effect of most heart diseases on myocardial T_1 has been investigated, mostly in single center studies. Select pathologies have been studied in large cohorts or multi-center studies, including studies on amyloidosis (147) and Anderson-Fabry disease (148). Additionally, cardiac T_1 mapping has been adopted in multiple national cohorts, including the UK biobank protocol and the German national cohort (149,150). These studies are some of the largest ongoing MRI projects to date. Following the clinical success demonstrated in the literature, cardiac T_1 mapping was adopted in disease specific clinical guidelines (151). Further increases in clinical integration and use in a growing number of cardiac MRI protocols are likely.

RELATED INITIATIVES, CHALLENGES AND OPPORTUNITIES

As described above, substantial efforts are needed for the development, validation and dissemination of quantitative MR techniques. These efforts require collaboration between technical researchers, translational researchers and clinicians, industry, and initiatives and institutions dedicated to the regulation and guidance of quantitative imaging measurements. Such initiatives and institutions include authorities for standardization of measurements such as Italy's Istituto Nazionale di Ricerca Metrologica (INRIM), the Korea Research Institute of Science and Standards (KRISS), the U.S.' National Institute of Standards and Technology (NIST), the UK's National Physical Laboratory (NPL), Germany's Physikalisch-Technische Bundesanstalt (PTB), and for the advancement of the development and use of imaging biomarkers, as performed by QIBA, the US National Cancer Institute through the Quantitative Imaging Network (QIN), the European Imaging Biomarkers Alliance (EIBALL), Japan-QIBA, the European Society of Radiology (ESR), and the ISMRM. Specifically, the major goal of the ISMRM Quantitative MR Study Group is to promulgate documentary and measurement standards for qMR methods in collaboration with national metrology institutes, academic and clinical MR sites, and through collaboration with existing study groups. Further, ongoing qMR improvements occur in the context of broad efforts to evaluate and optimize the value of MRI in medicine (152). Ultimately,

efforts to develop qMR methods should have broad value in medicine across countries and populations, beyond specialized research centers.

In addition, the development and validation of qMR methods is closely connected to the improvement of the reproducibility of MR research itself. There are multiple existing and emerging initiatives in this area, including the ISMRM Reproducible Research Study Group, and reproducibility has recently been emphasized by major journals such as *Magnetic Resonance in Medicine,* or the *Journal of Magnetic Resonance Imaging*. A related set of benchmarks for validation of quantitative imaging tools has been described by the Quantitative Imaging Network of the US' National Cancer Institute (153). Importantly, multiple consensus efforts and community challenges have emerged in recent years for specific qMR methods or applications, as well as for general optimization of qMR (6,20,37,142,154-163). Data standards are also essential for reproducibility and interoperability, making it easier to create transparent qMR workflows. Two standards that are relevant for qMR are the ISMRM-Raw Data format (https://ismrmrd.github.io/apidocs/1.5.0/), and the Brain Imaging Data Structure (BIDS) extension proposal for quantitative MRI (https://github.com/bids-standard/bids-specification/pull/508).

Further, various software packages developed, maintained and used by the community enable improved reproducibility by standardizing data processing pipelines. Supporting Information Figure S6 gives an overview of the user base of publicly available software with applications in qMR. Examples include various toolboxes hosted on the Matlab Central File exchange (https://www.mathworks.com/matlabcentral/fileexchange/), FSLTools (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslTools), OsiriX or Horos plugins, ImageJ (https://imagej.nih.gov/ij/), Bay Area Reconstruction Toolbox (BART, https://mrirecon.github.io/bart/), qMRlab (https://qmrlab.org) (164), Gadgetron (http://gadgetron.github.io/), Quantitative Imaging Tools (https://github.com/spinicist/QUIT), Michigan Image Reconstruction Toolbox (MIRT, http://github.com/JeffFessler/MIRT.jl), hMRI (https://hmri-group.github.io/hMRI-toolbox/), , QMRI tools (https://community.wolfram.com/groups/-/m/t/1661539), LCModel (http://s-provencher.com/lcmodel.shtml), Total Mapping Toolbox (TOMATO) (https://mrkonrad.github.io/TOMATO/html), QMRI Tools (https://mfroeling.github.io/QMRITools/), and others (e.g., vendor proprietary software and in-house or personal code). RSNA QIBA and NIBIB have also sponsored the development of digital

reference objects (DROs), which enable the testing of analysis tools to assess their bias and precision when working with quantitative data obtained with different acquisition parameters and varying levels of SNR. Example DROs for DCE-MRI and DWI are available from QIBA (https://qidw.rsna.org).

The transformation of MR into a truly quantitative diagnostic modality has enormous potential to impact research and clinical care. However, the development, validation, and dissemination of quantitative MR methods is faced with multiple challenges, particularly the complexity and cost of the required validation studies, as highlighted by the above networks and initiatives. These challenges reinforce the need for collaboration between technical MR researchers, academic radiologists, and other clinicians, as well as industry, such as Original Equipment Manufacturers (OEMs) - vendors of MR systems and other MR equipment and software, pharmaceutical companies, contract research organizations, and others.

Finally, substantial recent efforts from the qMR community have focused on rapid multi-parametric mapping and machine learning (ML). Emerging multi-parametric mapping methods such as MR fingerprinting (165) and multi-tasking (81) enable quantitative mapping of several parameters with short scan times. These methods are highly promising for a variety of applications, and require careful development and validation to address a large number of potential confounding factors. ML methods, including radiomics and deep learning, have recently gained enormous interest in the field. Indeed, ML may contribute to different stages of the qMR pipeline, including image prescription, acquisition, reconstruction, post-processing, measurement, and analysis. Despite the potential impact of these methods, rigorous development and validation of ML-enabled qMR is needed. This development and validation pose new challenges and opportunities for ML-enabled qMR, including how to quantify and address confounding factors to achieve low bias and high reproducibility across patients, sites, and vendors, in much the same way as the more ‘traditional’ qMR methods highlighted in the present manuscript.

SUMMARY AND CONCLUSION

On behalf of the International Society for Magnetic Resonance in Medicine (ISMRM) Quantitative MR Study Group, this manuscript describes a framework for the development and validation of quantitative MR methods. With a focus on technical performance metrics (bias and precision), this

framework highlights the challenges as well as the research opportunities associated with quantitative MR methods. Overall, rigorous development and validation are critical components of the transformation of MR into a truly quantitative diagnostic modality. A summary of concluding recommendations to achieve this aim is provided in Table 3. Upon successful implementation of qMR methods, as well as clinical qualification of qMR-based biomarkers, qMR has the potential to substantially advance imaging in clinical applications and clinical research, and build a cornerstone of precision radiology.

Table 3: Recommendation for qMR development, validation and application.

Definitions	The measurand of interest needs to be clearly defined. How does the targeted measurand relate to other physical properties? For example, if a coefficient is determined, what is the reference quantity (e.g., MR-visible protons)?
Choice of pulse sequence	Select pulse sequences and parameters such that the measurand can be determined with low bias and high precision, subject to a set of timing, hardware, and other constraints. Acquisition design will often begin by selecting a pulse sequence where the measurand of interest can be directly probed, while minimizing the effect of confounding factors.
Choice of models	Proper biophysical modeling is difficult, but may avoid the pitfalls of various signal representations. Indeed, various models can often fit the data, but models that are not grounded on specific tissue assumptions are often more difficult to validate, and are also likely to suffer from poor reproducibility.
Rigorous validation	It is critical to perform systematic validation of the technical performance of emerging qMR methods. Importantly, even though early-stage validation is often focused on bias, evaluation of precision (repeatability and reproducibility) is essential to enable further clinical qualification and dissemination.
Structured evaluation	Well-structured reporting of the validation is an essential component of establishing a qMR method. The standard metrics being evaluated should be described clearly as discussed in the section "Technical Performance of qMR Methods" above. Future work from the community may establish a standardized structure for the Methods and Results sections of qMR manuscripts.

Real-world validation	Even at the stage of technical validation, it is important to evaluate the performance of qMR methods under conditions that are relevant to the real-world clinical environment. For example, performance may depend on the hardware available at different sites (e.g., academic vs non-academic). In addition, technical validation in a relevant patient population helps pave the way for subsequent clinical qualification and application.
Focus on reproducibility	Optimization and characterization of reproducibility across acquisition protocols, field strength, vendor, platform, etc, is critical in qMR. Indeed, in qualitative MR methods development, one is often interested in finding the optimal set of acquisition and processing parameters to maximize imaging performance (e.g., resolution, SNR). Although this optimal set of parameters is also relevant in qMR, the development of qMR methods that are reproducible across variations in the acquisition parameters is arguably even more important than the identification of the optimal parameters. This way, qMR methods are best suited for widespread dissemination across sites that may not be able to implement exactly optimized acquisitions.
Reproducibility vs. standardization	Certain qMR methods are highly reproducible across variations in acquisition parameters (within a certain range). For example, this is the case for PDFF measurement in the liver: when correcting for all relevant confounders, PDFF measurement is highly reproducible across field strength, echo time combinations, spatial resolution, and various other acquisition parameters. However, other qMR methods have poorer reproducibility, and their widespread dissemination would benefit highly from standardization of acquisitions (as well as processing) across sites and systems, as well as harmonization (see below).
Harmonization	Quantitative MR can benefit from harmonized acquisitions and tools. For example, standardized reference objects and tool validation methods, such as the use of DROs, provide common ground for comparison of imaging protocols across sites, vendors, and software analysis packages.
Realistic time-horizon	Development, validation, qualification, and dissemination of a qMR method is a slow, iterative process that may take more than a decade.
Consider the end goal	In qMR, the end goal is often to enable improved diagnosis, staging, and/or treatment monitoring of disease and generate

	increased value in the clinical work up. This goal is generally relevant even for technically focused researchers.
Mind the translational gap	Establishing new quantitative imaging biomarkers in actual clinical use is a lengthy process and requires many steps that may not be amenable to funding by traditional science-oriented grants or tenders. This may require creativity about the path to clinical integration and real-world use.
Clinical qualification is key	Clinical qualification is critical to achieve translation of a quantitative MR method to the clinic. This may be the most time-consuming step in the entire pipeline of qMR method development and evaluation.
Collaboration	Working together with stakeholders (technical, clinical, industrial) and across imaging modalities or scientific disciplines is critical. For example, accurate biophysical modelling will benefit from collaboration between clinical and preclinical MR scientists, but also between MR researchers and scientists studying tissues at smaller scales (e.g., cell cultures) or using different imaging technology (e.g., X-ray phase contrast imaging for tissue structure, near infrared spectroscopy for blood oxygenation properties, or microscopy). Collaboration with clinicians is of enormous value in qMR technique development, and helps create a virtuous loop of refinement of existing methods and conception of new methods that address existing clinical needs. Further, early-stage discussion and cooperation with industry is especially relevant, since CE/FDA-labelling is mandatory for clinical translation. A technique without labelling will not be widely adopted in clinical practice due to ethical concerns and regulatory issues.

ACKNOWLEDGMENTS

The authors would like to thank the ISMRM Quantitative MR Study Group Consortium, including those who responded to the online poll of the ISMRM Quantitative MR Study Group, as well as those who provided feedback during the public presentation on June 23, 2021, for their thoughtful feedback, which has led to an improved manuscript. We also acknowledge the efforts of the endorsers, as listed in the Supporting Information Table S2. We also thank the ISMRM Publications Committee for reviewing the work and coordinating the ISMRM Board of Trustees approval process.

CONFLICT OF INTEREST

Nancy Obuchowski is a paid consultant for RSNA's Quantitative Imaging Biomarker Alliance. Bettina Baessler is co-founder of Lernrad GmbH, Germany. Xavier Golay is a co-founder, CEO and shareholder of Gold Standard Phantoms. Diego Hernando is co-founder of Calimetrix, LLC.

REFERENCES

1. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69(3):89-95.
2. Abramson RG, Burton KR, Yu JP, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Lenchik L, Subramaniam RM. Methods and challenges in quantitative imaging biomarker development. Acad Radiol 2015;22(1):25-32.
3. Cercignani M, Dowell N, Tofts P, editors. Quantitative MRI of the brain: principles of Physical measurement. 2nd Edition ed. Boca Raton, FL, USA: CRC Press; 2018.
4. Cui Y, Zhang XP, Sun YS, Tang L, Shen L. Apparent diffusion coefficient: potential imaging biomarker for prediction and early detection of response to chemotherapy in hepatic metastases. Radiology 2008;248(3):894-900.
5. Abramson RG, Arlinghaus LR, Dula AN, Quarles CC, Stokes AM, Weis JA, Whisenant JG, Chekmenev EY, Zhukov I, Williams JM, Yankeelov TE. MR Imaging Biomarkers in Oncology Clinical Trials. Magn Reson Imaging Clin N Am 2016;24(1):11-29.
6. deSouza NM, Achten E, Alberich-Bayarri A, Bamberg F, Boellaard R, Clement O, Fournier L, Gallagher F, Golay X, Heussel CP, Jackson EF, Manniesing R, Mayerhofer ME, Neri E, O'Connor J, Oguz KK, Persson A, Smits M, van Beek EJR, Zech CJ, European Society of R. 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):87.
7. Modell B, Khan M, Darlison M, Westwood MA, Ingram D, Pennell DJ. Improved survival of thalassaemia major in the UK and relation to T2* cardiovascular magnetic resonance. J Cardiovasc Magn Reson 2008;10(1):42.
8. Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. Magn Reson Imaging Clin N Am 2010;18(3):337-357, ix.
9. Dixon WT. Simple proton spectroscopic imaging. Radiology 1984;153(1):189-194.
10. Reeder SB, Wen Z, Yu H, Pineda AR, Gold GE, Markl M, Pelc NJ. Multicoil Dixon chemical species separation with an iterative least-squares estimation method. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2004;51(1):35-45.
11. Pineda AR, Reeder SB, Wen Z, Pelc NJ. Cramer-Rao bounds for three-point decomposition of water and fat. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2005;54(3):625-635.
12. Liu CY, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: correction of bias from T(1) and noise. Magn Reson Med 2007;58(2):354-364.
13. Yu H, McKenzie CA, Shimakawa A, Vu AT, Brau AC, Beatty PJ, Pineda AR, Brittain JH, Reeder SB. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. Journal of magnetic resonance imaging : JMRI 2007;26(4):1153-1161.
14. O'Regan DP, Callaghan MF, Wylezinska-Arridge M, Fitzpatrick J, Naoumova RP, Hajnal JV, Schmitz SA. Liver fat content and T2*: simultaneous measurement by using breath-hold multiecho MR imaging at 3.0 T--feasibility. Radiology 2008;247(2):550-557.

15. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2008;60(5):1122-1134.
16. Hamilton G, Middleton MS, Bydder M, Yokoo T, Schwimmer JB, Kono Y, Patton HM, Lavine JE, Sirlin CB. Effect of PRESS and STEAM sequences on magnetic resonance spectroscopic liver fat quantification. J Magn Reson Imaging 2009;30(1):145-152.
17. Meisamy S, Hines CD, Hamilton G, Sirlin CB, McKenzie CA, Yu H, Brittain JH, Reeder SB. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology 2011;258(3):767-775.
18. Schwarzbauer C, Syha J, Haase A. Quantification of regional blood volumes by rapid T1 mapping. Magn Reson Med 1993;29(5):709-712.
19. Everett RJ, Stirrat CG, Semple SI, Newby DE, Dweck MR, Mirsadraee S. Assessment of myocardial fibrosis with T1 mapping MRI. Clin Radiol 2016;71(8):768-778.
20. Messroghli DR, Moon JC, Ferreira VM, Grosse-Wortmann L, He T, Kellman P, Mascherbauer J, Nezafat R, Salerno M, Schelbert EB, Taylor AJ, Thompson R, Ugander M, van Heeswijk RB, Friedrich MG. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). J Cardiovasc Magn Reson 2017;19(1):75.
21. Mewton N, Liu CY, Croisille P, Bluemke D, Lima JA. Assessment of myocardial fibrosis with cardiovascular magnetic resonance. J Am Coll Cardiol 2011;57(8):891-903.
22. Amano Y, Takeda M, Tachi M, Kitamura M, Kumita S. Myocardial fibrosis evaluated by Look-Locker and late gadolinium enhancement magnetic resonance imaging in apical hypertrophic cardiomyopathy: association with ventricular tachyarrhythmia and risk factors. J Magn Reson Imaging 2014;40(2):407-412.
23. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. Magn Reson Med 2004;52(1):141-146.
24. Piechnik SK, Ferreira VM, Dall'Armellina E, Cochlin LE, Greiser A, Neubauer S, Robson MD. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. J Cardiovasc Magn Reson 2010;12:69.
25. Radenkovic D, Weingartner S, Ricketts L, Moon JC, Captur G. T1 mapping in cardiac MRI. Heart Fail Rev 2017;22(4):415-430.
26. Higgins DM, Moon JC. Review of T1 Mapping Methods: Comparative Effectiveness Including Reproducibility Issues. Current Cardiovascular Imaging Reports 2014;7(3):9252.
27. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, Guimaraes AR, Filice R, Zhang Z, Sullivan DC, Group QTW. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 2015;24(1):9-26.

28. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, Kondratovich M, McShane LM, Reeves AP, Barboriak DP, Guimaraes AR, Wahl RL, Group R-QMW. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015;277(3):813-825.
29. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gonen M, Zahlmann G, Kondratovich MV, O'Donnell K, Petrick N, Cole PE, Garra B, Sullivan DC, Group QTPW. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27-67.
30. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, Barnhart HX, Jackson EF, Giger ML, Pennello G, Toledano AY, Kalpathy-Cramer J, Apanasovich TV, Kinahan PE, Myers KJ, Goldgof DB, Barboriak DP, Gillies RJ, Schwartz LH, Sullivan DC, Algorithm Comparison Working G. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. Stat Methods Med Res 2015;24(1):68-106.
31. Keenan KE, Ainslie M, Barker AJ, Boss MA, Cecil KM, Charles C, Chenevert TL, Clarke L, Evelhoch JL, Finn P, Gembris D, Gunter JL, Hill DLG, Jack CR, Jr., Jackson EF, Liu G, Russek SE, Sharma SD, Steckner M, Stupic KF, Trzasko JD, Yuan C, Zheng J. Quantitative magnetic resonance imaging phantoms: A review and the need for a system phantom. Magn Reson Med 2018;79(1):48-61.
32. Boss M, Dienstfrey A, Gimbutas Z, Keenan K, Splett J, Stupic K, Russek S. Magnetic Resonance Imaging Biomarker Calibration Service: Proton Spin Relaxation Times. Gaithersburg, MD: National Institute of Standards and Technology; 2018.
33. Stupic KF, Ainslie M, Boss MA, Charles C, Dienstfrey AM, Evelhoch JL, Finn P, Gimbutas Z, Gunter JL, Hill DLG, Jack CR, Jackson EF, Karaulanov T, Keenan KE, Liu G, Martin MN, Prasad PV, Rentz NS, Yuan C, Russek SE. A standard system phantom for magnetic resonance imaging. Magn Reson Med 2021.
34. Marques JP, Meineke J, Milovic C, Bilgic B, Chan KS, Hedouin R, van der Zwaag W, Langkammer C, Schweser F. QSM reconstruction challenge 2.0: A realistic in silico head phantom for MRI data simulation and evaluation of susceptibility mapping procedures. Magn Reson Med 2021;86(1):526-542.
35. Obuchowski NA, Buckler A, Kinahan P, Chen-Mayer H, Petrick N, Barboriak DP, Bullen J, Barnhart H, Sullivan DC. Statistical Issues in Testing Conformance with the Quantitative Imaging Biomarker Alliance (QIBA) Profile Claims. Acad Radiol 2016;23(4):496-506.
36. Obuchowski NA, Bullen J. Quantitative imaging biomarkers: effect of sample size and bias on confidence interval coverage. Stat Methods Med Res 2018;27(10):3139-3150.
37. Captur G, Bhandari A, Bruhl R, Ittermann B, Keenan KE, Yang Y, Eames RJ, Benedetti G, Torlasco C, Ricketts L, Boubertakh R, Fatih N, Greenwood JP, Paulis LEM, Lawton CB, Bucciarelli-Ducci C, Lamb HJ, Steeds R, Leung SW, Berry C, Valentin S, Flett A, de Lange C, DeCobelli F, Viallon M, Croisille P, Higgins DM, Greiser A, Pang W, Hamilton-Craig C, Strugnell WE, Dresselaers T, Barison A, Dawson D, Taylor AJ, Mongeon FP, Plein S, Messroghli D, Al-Mallah M, Grieve SM, Lombardi M, Jang J, Salerno M, Chaturvedi N, Kellman P, Bluemke DA, Nezafat R, Gatehouse P, Moon JC, Consortium TM. T1 mapping performance and measurement repeatability: results from

the multi-national T1 mapping standardization phantom program (T1MES). J Cardiovasc Magn Reson 2020;22(1):31.
38. Hernando D, Kuhn JP, Mensel B, Volzke H, Puls R, Hosten N, Reeder SB. R2* estimation using "in-phase" echoes in the presence of fat: the effects of complex spectrum of fat. J Magn Reson Imaging 2013;37(3):717-726.
39. Obuchowski NA, Mozley PD, Matthews D, Buckler A, Bullen J, Jackson E. Statistical Considerations for Planning Clinical Trials with Quantitative Imaging Biomarkers. J Natl Cancer Inst 2019;111(1):19-26.
40. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, Huang W, Noworolski SM, Young RJ, Shiroishi MS. 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(7):e101-e121.
41. Huang EP, Wang X-F, Choudhury KR, McShane LM, Gönen M, Ye J, Buckler AJ, Kinahan PE, Reeves AP, Jackson EF. Meta-analysis of the technical performance of an imaging procedure: guidelines and statistical methodology. Stat Methods Med Res 2015;24(1):141-174.
42. Obuchowski NA, Barnhart HX, Buckler AJ, Pennello G, Wang X-F, Kalpathy-Cramer J, Kim HJ, Reeves AP, Group CEW. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. Stat Methods Med Res 2015;24(1):107-140.
43. RSNA. QIBA Profile: Diffusion-Weighted Magnetic Resonance Imaging (DWI). 2019.
44. Hernando D, Cook RJ, Qazi N, Longhurst CA, Diamond CA, Reeder SB. Complex confounder-corrected R2* mapping for liver iron quantification with MRI. Eur Radiol 2021;31(1):264-275.
45. Hernando D, Zhao R, Mattison R, Reeder S, Yokoo T, Pedrosa I, Yuan Q, Karampinos D, Ruschke S, Kamel I, Aliyari M, Zhong X, Vasanawala S, Taviani V. Multi-center, multi-vendor reproducibility of confounder-corrected R2* mapping for liver iron quantification at 1.5T and 3T: interim results. 2019 March 17, 2019; Orlando, FL. Society of Abdominal Imaging Annual Meeting.
46. Hernando D, Sharma SD, Aliyari Ghasabeh M, Alvis BD, Arora SS, Hamilton G, Pan L, Shaffer JM, Sofue K, Szeverenyi NM, Welch EB, Yuan Q, Bashir MR, Kamel IR, Rice MJ, Sirlin CB, Yokoo T, Reeder SB. Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat-water phantom. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2017;77(4):1516-1524.
47. Hu HH, Yokoo T, Bashir MR, Sirlin CB, Hernando D, Malyarenko D, Chenevert TL, Smith MA, Serai SD, Middleton MS, Henderson WC, Hamilton G, Shaffer J, Shu Y, Tkach JA, Trout AT, Obuchowski N, Brittain JH, Jackson EF, Reeder SB, Committee RQIBAPB. Linearity and Bias of Proton Density Fat Fraction as a Quantitative Imaging Biomarker: A Multicenter, Multiplatform, Multivendor Phantom Study. Radiology 2021;298(3):640-651.
48. Yokoo T, Serai SD, Pirasteh A, Bashir MR, Hamilton G, Hernando D, Hu HH, Hetterich H, Kuhn JP, Kukuk GM, Loomba R, Middleton MS, Obuchowski NA, Song JS, Tang A, Wu X, Reeder SB, Sirlin CB, Committee R-QPB. Linearity, Bias, and Precision of

Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis. Radiology 2018;286(2):486-498.
49. Bannas P, Kramer H, Hernando D, Agni R, Cunningham AM, Mandal R, Motosugi U, Sharma SD, Munoz del Rio A, Fernandez L, Reeder SB. Quantitative magnetic resonance imaging of hepatic steatosis: Validation in ex vivo human livers. Hepatology 2015;62(5):1444-1455.
50. Hines CD, Frydrychowicz A, Hamilton G, Tudorascu DL, Vigen KK, Yu H, McKenzie CA, Sirlin CB, Brittain JH, Reeder SB. T(1) independent, T(2) (*) corrected chemical shift based fat-water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. Journal of magnetic resonance imaging : JMRI 2011;33(4):873-881.
51. Serai SD, Dillman JR, Trout AT. Proton Density Fat Fraction Measurements at 1.5- and 3-T Hepatic MR Imaging: Same-Day Agreement among Readers and across Two Imager Manufacturers. Radiology 2017;284(1):244-254.
52. Roujol S, Weingartner S, Foppa M, Chow K, Kawaji K, Ngo LH, Kellman P, Manning WJ, Thompson RB, Nezafat R. Accuracy, precision, and reproducibility of four T1 mapping sequences: a head-to-head comparison of MOLLI, ShMOLLI, SASHA, and SAPPHIRE. Radiology 2014;272(3):683-689.
53. Weingartner S, Messner NM, Budjan J, Lossnitzer D, Mattler U, Papavassiliu T, Zollner FG, Schad LR. Myocardial T1-mapping at 3T using saturation-recovery: reference values, precision and comparison with MOLLI. J Cardiovasc Magn Reson 2016;18(1):84.
54. Fontana M, White SK, Banypersad SM, Sado DM, Maestrini V, Flett AS, Piechnik SK, Neubauer S, Roberts N, Moon JC. Comparison of T1 mapping techniques for ECV quantification. Histological validation and reproducibility of ShMOLLI versus multibreath-hold T1 quantification equilibrium contrast CMR. J Cardiovasc Magn Reson 2012;14:88.
55. Kellman P, Hansen MS. T1-mapping in the heart: accuracy and precision. J Cardiovasc Magn Reson 2014;16:2.
56. Graham-Brown MP, Rutherford E, Levelt E, March DS, Churchward DR, Stensel DJ, McComb C, Mangion K, Cockburn S, Berry C, Moon JC, Mark PB, Burton JO, McCann GP. Native T1 mapping: inter-study, inter-observer and inter-center reproducibility in hemodialysis patients. J Cardiovasc Magn Reson 2017;19(1):21.
57. Chow K, Flewitt JA, Green JD, Pagano JJ, Friedrich MG, Thompson RB. Saturation recovery single-shot acquisition (SASHA) for myocardial T(1) mapping. Magn Reson Med 2014;71(6):2082-2095.
58. Raman FS, Kawel-Boehm N, Gai N, Freed M, Han J, Liu CY, Lima JA, Bluemke DA, Liu S. Modified look-locker inversion recovery T1 mapping indices: assessment of accuracy and reproducibility between magnetic resonance scanners. J Cardiovasc Magn Reson 2013;15:64.
59. Dabir D, Child N, Kalra A, Rogers T, Gebker R, Jabbour A, Plein S, Yu CY, Otton J, Kidambi A, McDiarmid A, Broadbent D, Higgins DM, Schnackenburg B, Foote L, Cummins C, Nagel E, Puntmann VO. Reference values for healthy human myocardium using a T1 mapping methodology: results from the International T1 Multicenter cardiovascular magnetic resonance study. J Cardiovasc Magn Reson 2014;16:69.

60. Popescu IA, Werys K, Zhang Q, Puchta H, Hann E, Lukaschuk E, Ferreira VM, Piechnik SK. Standardization of T1-mapping in cardiovascular magnetic resonance using clustered structuring for benchmarking normal ranges. Int J Cardiol 2021;326:220-225.
61. Glover GH, Schneider E. Three-point Dixon technique for true water/fat decomposition with B0 inhomogeneity correction. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 1991;18(2):371-383.
62. Malyarenko DI, Ross BD, Chenevert TL. Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements. Magn Reson Med 2014;71(3):1312-1323.
63. Bernstein MA, Zhou XJ, Polzin JA, King KF, Ganin A, Pelc NJ, Glover GH. Concomitant gradient terms in phase contrast MR: analysis and correction. Magn Reson Med 1998;39(2):300-308.
64. Colgan TJ, Hernando D, Sharma SD, Reeder SB. The effects of concomitant gradients on chemical shift encoded MRI. Magn Reson Med 2017;78(2):730-738.
65. Jezzard P, Barnett AS, Pierpaoli C. Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. Magn Reson Med 1998;39(5):801-812.
66. Benner T, van der Kouwe AJ, Kirsch JE, Sorensen AG. Real-time RF pulse adjustment for B0 drift correction. Magn Reson Med 2006;56(1):204-209.
67. Peters DC, Derbyshire JA, McVeigh ER. Centering the projection reconstruction trajectory: reducing gradient delay errors. Magn Reson Med 2003;50(1):1-6.
68. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: A complex problem with many partial solutions. J Magn Reson Imaging 2015;42(4):887-901.
69. Novikov DS, Kiselev VG, Jespersen SN. On modeling. Magn Reson Med 2018;79(6):3172-3193.
70. McRobbie D, Lerski R, Straughan K. Slice profile effects and their calibration and correction in quantitative NMR imaging. Physics in Medicine & Biology 1987;32(8):971.
71. Malik SJ, Kenny GD, Hajnal JV. Slice profile correction for transmit sensitivity mapping using actual flip angle imaging. Magn Reson Med 2011;65(5):1393-1399.
72. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 1995;34(6):910-914.
73. Jones DK, Basser PJ. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn Reson Med 2004;52(5):979-993.
74. Hernando D, Kramer JH, Reeder SB. Multipeak fat-corrected complex R2* relaxometry: theory, optimization, and clinical validation. Magn Reson Med 2013;70(5):1319-1331.
75. Bydder M, Yokoo T, Hamilton G, Middleton MS, Chavez AD, Schwimmer JB, Lavine JE, Sirlin CB. Relaxation effects in the quantification of fat using gradient echo imaging. Magnetic resonance imaging 2008;26(3):347-359.
76. Horng DE, Hernando D, Reeder SB. Quantification of liver fat in the presence of iron overload. J Magn Reson Imaging 2017;45(2):428-439.
77. Hamilton G, Yokoo T, Bydder M, Cruite I, Schroeder ME, Sirlin CB, Middleton MS. In vivo characterization of the liver fat (1)H MR spectrum. NMR in biomedicine 2011;24(7):784-790.
78. Yu H, Shimakawa A, Hines CD, McKenzie CA, Hamilton G, Sirlin CB, Brittain JH, Reeder SB. Combination of complex-based and magnitude-based multiecho water-fat

separation for accurate quantification of fat-fraction. Magn Reson Med 2011;66(1):199-206.
79. Hernando D, Hines CD, Yu H, Reeder SB. Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2012;67(3):638-644.
80. Weingartner S, Akcakaya M, Basha T, Kissinger KV, Goddu B, Berg S, Manning WJ, Nezafat R. Combined saturation/inversion recovery sequences for improved evaluation of scar and diffuse fibrosis in patients with arrhythmia or heart rate variability. Magn Reson Med 2014;71(3):1024-1034.
81. Christodoulou AG, Shaw JL, Nguyen C, Yang Q, Xie Y, Wang N, Li D. Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging. Nat Biomed Eng 2018;2(4):215-226.
82. Weingartner S, Shenoy C, Rieger B, Schad LR, Schulz-Menger J, Akcakaya M. Temporally resolved parametric assessment of Z-magnetization recovery (TOPAZ): Dynamic myocardial T1 mapping using a cine steady-state look-locker approach. Magn Reson Med 2018;79(4):2087-2100.
83. Guo R, Cai X, Kucukseymen S, Rodriguez J, Paskavitz A, Pierce P, Goddu B, Thompson RB, Nezafat R. Free-breathing simultaneous myocardial T1 and T2 mapping with whole left ventricle coverage. Magn Reson Med 2021;85(3):1308-1321.
84. Robson MD, Piechnik SK, Tunnicliffe EM, Neubauer S. T1 measurements in the human myocardium: the effects of magnetization transfer on the SASHA and MOLLI sequences. Magn Reson Med 2013;70(3):664-670.
85. Kellman P, Herzka DA, Arai AE, Hansen MS. Influence of Off-resonance in myocardial T1-mapping using SSFP based MOLLI method. J Cardiovasc Magn Reson 2013;15:63.
86. Cooper MA, Nguyen TD, Spincemaille P, Prince MR, Weinsaft JW, Wang Y. How Accurate Is MOLLI T1 Mapping In Vivo? Validation by Spin Echo Methods. PLOS ONE 2014;9(9):e107327.
87. Cameron D, Vassiliou VS, Higgins DM, Gatehouse PD. Towards accurate and precise T 1 and extracellular volume mapping in the myocardium: a guide to current pitfalls and their solutions. MAGMA 2018;31(1):143-163.
88. Weingartner S, Messner NM, Zollner FG, Akcakaya M, Schad LR. Black-blood native T1 mapping: Blood signal suppression for reduced partial voluming in the myocardium. Magn Reson Med 2017;78(2):484-493.
89. Motosugi U, Hernando D, Wiens C, Bannas P, Reeder SB. High SNR Acquisitions Improve the Repeatability of Liver Fat Quantification Using Confounder-corrected Chemical Shift-encoded MR Imaging. Magn Reson Med Sci 2017;16(4):332-339.
90. Boudreau M, Pike GB. Sensitivity regularization of the Cramér-Rao lower bound to minimize B1 nonuniformity effects in quantitative magnetization transfer imaging. Magn Reson Med 2018;80(6):2560-2572.
91. Drakesmith M, Harms R, Rudrapatna SU, Parker GD, Evans CJ, Jones DK. Estimating axon conduction velocity in vivo from microstructural MRI. Neuroimage 2019;203:116186.
92. Scharf LL, LT M. Geometry of the Cramér-Rao Bound. In: Proceedings of the IEEE Sixth SP Workshop on Statistical Signal and Array Processing. Victoria, BC, Canada1992. p 5-8.

93. Akcakaya M, Weingartner S, Roujol S, Nezafat R. On the selection of sampling points for myocardial T1 mapping. Magn Reson Med 2015;73(5):1741-1753.
94. Kellman P, Xue H, Chow K, Spottiswoode BS, Arai AE, Thompson RB. Optimized saturation recovery protocols for T1-mapping in the heart: influence of sampling strategies on precision. J Cardiovasc Magn Reson 2014;16:55.
95. Jang J, Bellm S, Roujol S, Basha TA, Nezafat M, Kato S, Weingartner S, Nezafat R. Comparison of spoiled gradient echo and steady-state free-precession imaging for native myocardial T1 mapping using the slice-interleaved T1 mapping (STONE) sequence. NMR Biomed 2016;29(10):1486-1496.
96. Shaw JL, Yang Q, Zhou Z, Deng Z, Nguyen C, Li D, Christodoulou AG. Free-breathing, non-ECG, continuous myocardial T1 mapping with cardiovascular magnetic resonance multitasking. Magn Reson Med 2019;81(4):2450-2463.
97. Qi H, Jaubert O, Bustin A, Cruz G, Chen H, Botnar R, Prieto C. Free-running 3D whole heart myocardial T1 mapping with isotropic spatial resolution. Magn Reson Med 2019;82(4):1331-1342.
98. Chow K, Yang Y, Shaw P, Kramer CM, Salerno M. Robust free-breathing SASHA T1 mapping with high-contrast image registration. J Cardiovasc Magn Reson 2016;18(1):47.
99. Weingartner S, Roujol S, Akcakaya M, Basha TA, Nezafat R. Free-breathing multislice native myocardial T1 mapping using the slice-interleaved T1 (STONE) sequence. Magn Reson Med 2015;74(1):115-124.
100. Weingartner S, Akcakaya M, Roujol S, Basha T, Stehning C, Kissinger KV, Goddu B, Berg S, Manning WJ, Nezafat R. Free-breathing post-contrast three-dimensional T1 mapping: Volumetric assessment of myocardial T1 values. Magn Reson Med 2015;73(1):214-222.
101. Nordio G, Bustin A, Henningsson M, Rashid I, Chiribiri A, Ismail T, Odille F, Prieto C, Botnar RM. 3D SASHA myocardial T1 mapping with high accuracy and improved precision. MAGMA 2019;32(2):281-289.
102. Wang Y, Liu T. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. Magn Reson Med 2015;73(1):82-101.
103. Hancu I, Liu J, Hua Y, Lee SK. Electrical properties tomography: Available contrast and reconstruction capabilities. Magn Reson Med 2019;81(2):803-810.
104. Liu J, Wang Y, Katscher U, He B. Electrical Properties Tomography Based on $B_{{1}}$ Maps in MRI: Principles, Applications, and Challenges. IEEE Trans Biomed Eng 2017;64(11):2515-2530.
105. Bydder M, Yokoo T, Yu H, Carl M, Reeder SB, Sirlin CB. Constraining the initial phase in water-fat separation. Magnetic resonance imaging 2011;29(2):216-221.
106. Horng DE, Hernando D, Hines CD, Reeder SB. Comparison of R2* correction methods for accurate fat quantification in fatty liver. J Magn Reson Imaging 2013;37(2):414-422.
107. Wang X, Hernando D, Reeder SB. Sensitivity of chemical shift-encoded fat quantification to calibration of fat MR spectrum. Magn Reson Med 2016;75(2):845-851.
108. Deichmann R, Haase A. Quantification of T1 values by SNAPSHOT-FLASH NMR imaging. Journal of Magnetic Resonance (1969) 1992;96(3):608-612.
109. Shao J, Rapacchi S, Nguyen KL, Hu P. Myocardial T1 mapping at 3.0 tesla using an inversion recovery spoiled gradient echo readout and bloch equation simulation with slice profile correction (BLESSPC) T1 estimation algorithm. J Magn Reson Imaging 2016;43(2):414-425.

110. Xanthis CG, Bidhult S, Kantasis G, Heiberg E, Arheden H, Aletras AH. Parallel simulations for QUAntifying RElaxation magnetic resonance constants (SQUAREMR): an example towards accurate MOLLI T1 measurements. J Cardiovasc Magn Reson 2015;17:104.
111. Veraart J, Sijbers J, Sunaert S, Leemans A, Jeurissen B. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. Neuroimage 2013;81:335-346.
112. Bertsekas DP. Nonlinear Programming: Athena Scientific; 2016.
113. Xue H, Greiser A, Zuehlsdorff S, Jolly MP, Guehring J, Arai AE, Kellman P. Phase-sensitive inversion recovery for myocardial T1 mapping with motion correction and parametric fitting. Magn Reson Med 2013;69(5):1408-1420.
114. Hunter DJ, Losina E, Guermazi A, Burstein D, Lassere MN, Kraus V. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. Curr Drug Targets 2010;11(5):536-545.
115. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood) 2018;243(3):213-221.
116. Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Abramson RG, Burton KR, John-Paul JY, Scalzetti EM, Yankeelov TE, Subramaniam RM. Clinical utility of quantitative imaging. Acad Radiol 2015;22(1):33-49.
117. Brisman JL, Pile-Spellman J, Konstas AA. Clinical utility of quantitative magnetic resonance angiography in the assessment of the underlying pathophysiology in a variety of cerebrovascular disorders. Eur J Radiol 2012;81(2):298-302.
118. Eskreis-Winkler S, Zhang Y, Zhang J, Liu Z, Dimov A, Gupta A, Wang Y. The clinical utility of QSM: disease diagnosis, medical management, and surgical planning. NMR Biomed 2017;30(4):e3668.
119. Gallagher EJ. Clinical utility of likelihood ratios. Ann Emerg Med 1998;31(3):391-397.
120. Dobbin KK, Cesano A, Alvarez J, Hawtin R, Janetzki S, Kirsch I, Masucci GV, Robbins PB, Selvan SR, Streicher HZ, Zhang J, Butterfield LH, Thurin M. Validation of biomarkers to predict response to immunotherapy in cancer: Volume II - clinical validation and regulatory considerations. J Immunother Cancer 2016;4:77.
121. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins D, Hammoud DA, Rustin GJ, Taouli B, Choyke PL. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 2009;11(2):102-125.
122. Baessler B, Schaarschmidt F, Dick A, Stehning C, Schnackenburg B, Michels G, Maintz D, Bunck AC. Mapping tissue inhomogeneity in acute myocarditis: a novel analytical approach to quantitative myocardial edema imaging by T2-mapping. J Cardiovasc Magn Reson 2015;17:115.
123. Baessler B, Schaarschmidt F, Stehning C, Schnackenburg B, Giolda A, Maintz D, Bunck AC. Reproducibility of three different cardiac T2 -mapping sequences at 1.5T. J Magn Reson Imaging 2016;44(5):1168-1178.
124. Gu J, Liu S, Du S, Zhang Q, Xiao J, Dong Q, Xin Y. Diagnostic value of MRI-PDFF for hepatic steatosis in patients with non-alcoholic fatty liver disease: a meta-analysis. Eur Radiol 2019;29(7):3564-3573.
125. Loomba R, Neuschwander-Tetri BA, Sanyal A, Chalasani N, Diehl AM, Terrault N, Kowdley K, Dasarathy S, Kleiner D, Behling C, Lavine J, Van Natta M, Middleton M,

Tonascia J, Sirlin C, Network NCR. Multicenter Validation of Association Between Decline in MRI-PDFF and Histologic Response in NASH. Hepatology 2020;72(4):1219-1229.
126. Stine JG, Munaganuru N, Barnard A, Wang JL, Kaulback K, Argo CK, Singh S, Fowler KJ, Sirlin CB, Loomba R. Change in MRI-PDFF and Histologic Response in Patients With Nonalcoholic Steatohepatitis: A Systematic Review and Meta-Analysis. Clin Gastroenterol Hepatol 2020.
127. Rehm JL, Wolfgram PM, Hernando D, Eickhoff JC, Allen DB, Reeder SB. Proton density fat-fraction is an accurate biomarker of hepatic steatosis in adolescent girls and young women. Eur Radiol 2015;25(10):2921-2930.
128. Puchner SB, Lu MT, Mayrhofer T, Liu T, Pursnani A, Ghoshhajra BB, Truong QA, Wiviott SD, Fleg JL, Hoffmann U, Ferencik M. High-risk coronary plaque at coronary CT angiography is associated with nonalcoholic fatty liver disease, independent of coronary plaque and stenosis burden: results from the ROMICAT II trial. Radiology 2015;274(3):693-701.
129. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67(1):328-357.
130. Faragli A, Merz S, Muzio FPL, Doeblin P, Tanacli R, Kolp C, Abawi D, Otvos J, Stehning C, Schnackenburg B, Pieske B, Post H, Klopfleisch R, Alogna A, Kelle S. Estimation of total collagen volume: a T1 mapping versus histological comparison study in healthy Landrace pigs. Int J Cardiovasc Imaging 2020;36(9):1761-1769.
131. Reiter U, Reiter C, Krauter C, Fuchsjager M, Reiter G. Cardiac magnetic resonance T1 mapping. Part 2: Diagnostic potential and applications. Eur J Radiol 2018;109:235-247.
132. Hamilton-Craig CR, Strudwick MW, Galloway GJ. T1 Mapping for Myocardial Fibrosis by Cardiac Magnetic Resonance Relaxometry-A Comprehensive Technical Review. Front Cardiovasc Med 2016;3:49.
133. Lurz JA, Luecke C, Lang D, Besler C, Rommel KP, Klingel K, Kandolf R, Adams V, Schöne K, Hindricks G, Schuler G, Linke A, Thiele H, Gutberlet M, Lurz P. CMR-Derived Extracellular Volume Fraction as a Marker for Myocardial Fibrosis: The Importance of Coexisting Myocardial Inflammation. JACC Cardiovasc Imaging 2018;11(1):38-45.
134. Fehrmann A, Treutlein M, Rudolph T, Rudolph V, Weiss K, Giese D, Bunck AC, Maintz D, Baessler B. Myocardial T1 and T2 mapping in severe aortic stenosis: Potential novel insights into the pathophysiology of myocardial remodelling. Eur J Radiol 2018;107:76-83.
135. Aherne E, Chow K, Carr J. Cardiac T1 mapping: Techniques and applications. J Magn Reson Imaging 2020;51(5):1336-1356.
136. Pan JA, Kerwin MJ, Salerno M. Native T1 Mapping, Extracellular Volume Mapping, and Late Gadolinium Enhancement in Cardiac Amyloidosis: A Meta-Analysis. JACC Cardiovasc Imaging 2020;13(6):1299-1310.
137. Duan C, Zhu Y, Jang J, Rodriguez J, Neisius U, Fahmy AS, Nezafat R. Non-contrast myocardial infarct scar assessment using a hybrid native T1 and magnetization transfer imaging sequence at 1.5T. Magn Reson Med 2019;81(5):3192-3201.

138. Lopez K, Neji R, Mukherjee RK, Whitaker J, Phinikaridou A, Razavi R, Prieto C, Roujol S, Botnar R. Contrast-free high-resolution 3D magnetization transfer imaging for simultaneous myocardial scar and cardiac vein visualization. MAGMA 2020;33(5):627-640.
139. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Faivre-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh DM, Lacombe D, Lambin P, Lassau N, Leach MO, Lee TY, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJ, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Stroobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14(3):169-186.
140. Tofts PS, Collins DJ. Multicentre imaging measurements for oncology and in the brain. Br J Radiol 2011;84 Spec No 2(Spec Iss 2):S213-226.
141. Captur G, Gatehouse P, Keenan KE, Heslinga FG, Bruehl R, Prothmann M, Graves MJ, Eames RJ, Torlasco C, Benedetti G, Donovan J, Ittermann B, Boubertakh R, Bathgate A, Royet C, Pang W, Nezafat R, Salerno M, Kellman P, Moon JC. A medical device-grade T1 and ECV phantom for global T1 mapping quality assurance-the T1 Mapping and ECV Standardization in cardiovascular magnetic resonance (T1MES) program. J Cardiovasc Magn Reson 2016;18(1):58.
142. Keenan KE, Biller JR, Delfino JG, Boss MA, Does MD, Evelhoch JL, Griswold MA, Gunter JL, Hinks RS, Hoffman SW, Kim G, Lattanzi R, Li X, Marinelli L, Metzger GJ, Mukherjee P, Nordstrom RJ, Peskin AP, Perez E, Russek SE, Sahiner B, Serkova N, Shukla-Dave A, Steckner M, Stupic KF, Wilmes LJ, Wu HH, Zhang H, Jackson EF, Sullivan DC. Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. J Magn Reson Imaging 2019;49(7):e26-e39.
143. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Aerts HJ, Bendriem B, Bendtsen C, Boellaard R, Boone JM, Cole PE, Conklin JJ, Dorfman GS, Douglas PS, Eidsaunet W, Elsinger C, Frank RA, Gatsonis C, Giger ML, Gupta SN, Gustafson D, Hoekstra OS, Jackson EF, Karam L, Kelloff GJ, Kinahan PE, McLennan G, Miller CG, Mozley PD, Muller KE, Patt R, Raunig D, Rosen M, Rupani H, Schwartz LH, Siegel BA, Sorensen AG, Wahl RL, Waterton JC, Wolf W, Zahlmann G, Zimmerman B. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 2011;259(3):875-884.
144. Radiological Society of North America Quantitative Imaging Biomarkers Alliance.
145. Kuhn JP, Meffert P, Heske C, Kromrey ML, Schmidt CO, Mensel B, Volzke H, Lerch MM, Hernando D, Mayerle J, Reeder SB. Prevalence of Fatty Liver Disease and Hepatic Iron Overload in a Northeastern German Population by Using Quantitative MR Imaging. Radiology 2017;284(3):706-716.
146. Ajmera V, Park CC, Caussy C, Singh S, Hernandez C, Bettencourt R, Hooker J, Sy E, Behling C, Xu R, Middleton MS, Valasek MA, Faulkner C, Rizo E, Richards L, Sirlin CB, Loomba R. Magnetic Resonance Imaging Proton Density Fat Fraction Associates

With Progression of Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2018;155(2):307-310 e302.
147. Baggiano A, Boldrini M, Martinez-Naharro A, Kotecha T, Petrie A, Rezk T, Gritti M, Quarta C, Knight DS, Wechalekar AD, Lachmann HJ, Perlini S, Pontone G, Moon JC, Kellman P, Gillmore JD, Hawkins PN, Fontana M. Noncontrast Magnetic Resonance for the Diagnosis of Cardiac Amyloidosis. JACC Cardiovasc Imaging 2020;13(1 Pt 1):69-80.
148. Augusto JB, Nordin S, Vijapurapu R, Baig S, Bulluck H, Castelletti S, Alfarih M, Knott K, Captur G, Kotecha T, Ramaswami U, Tchan M, Geberhiwot T, Fontana M, Steeds RP, Hughes D, Kozor R, Moon JC. Myocardial Edema, Myocyte Injury, and Disease Severity in Fabry Disease. Circ Cardiovasc Imaging 2020;13(3):e010171.
149. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, Young AA, Hudson S, Weale P, Garratt S, Collins R, Piechnik S, Neubauer S. UK Biobank's cardiovascular magnetic resonance protocol. J Cardiovasc Magn Reson 2016;18:8.
150. Bamberg F, Kauczor HU, Weckbach S, Schlett CL, Forsting M, Ladd SC, Greiser KH, Weber MA, Schulz-Menger J, Niendorf T, Pischon T, Caspers S, Amunts K, Berger K, Bulow R, Hosten N, Hegenscheid K, Kroncke T, Linseisen J, Gunther M, Hirsch JG, Kohn A, Hendel T, Wichmann HE, Schmidt B, Jockel KH, Hoffmann W, Kaaks R, Reiser MF, Volzke H, German National Cohort MRISI. Whole-Body MR Imaging in the German National Cohort: Rationale, Design, and Technical Background. Radiology 2015;277(1):206-220.
151. Ferreira VM, Schulz-Menger J, Holmvang G, Kramer CM, Carbone I, Sechtem U, Kindermann I, Gutberlet M, Cooper LT, Liu P, Friedrich MG. Cardiovascular Magnetic Resonance in Nonischemic Myocardial Inflammation: Expert Recommendations. J Am Coll Cardiol 2018;72(24):3158-3176.
152. van Beek EJR, Kuhl C, Anzai Y, Desmond P, Ehman RL, Gong Q, Gold G, Gulani V, Hall-Craggs M, Leiner T, Lim CCT, Pipe JG, Reeder S, Reinhold C, Smits M, Sodickson DK, Tempany C, Vargas HA, Wang M. Value of MRI in medicine: More than just another test? J Magn Reson Imaging 2019;49(7):e14-e25.
153. Farahani K, Tata D, Nordstrom RJ. QIN Benchmarks for Clinical Translation of Quantitative Imaging Tools. Tomography 2019;5(1):1-6.
154. Wilson M, Andronesi O, Barker PB, Bartha R, Bizzi A, Bolan PJ, Brindle KM, Choi IY, Cudalbu C, Dydak U, Emir UE, Gonzalez RG, Gruber S, Gruetter R, Gupta RK, Heerschap A, Henning A, Hetherington HP, Huppi PS, Hurd RE, Kantarci K, Kauppinen RA, Klomp DWJ, Kreis R, Kruiskamp MJ, Leach MO, Lin AP, Luijten PR, Marjanska M, Maudsley AA, Meyerhoff DJ, Mountford CE, Mullins PG, Murdoch JB, Nelson SJ, Noeske R, Oz G, Pan JW, Peet AC, Poptani H, Posse S, Ratai EM, Salibi N, Scheenen TWJ, Smith ICP, Soher BJ, Tkac I, Vigneron DB, Howe FA. Methodological consensus on clinical proton MRS of the brain: Review and recommendations. Magn Reson Med 2019;82(2):527-550.
155. Near J, Harris AD, Juchem C, Kreis R, Marjanska M, Oz G, Slotboom J, Wilson M, Gasparovic C. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. NMR Biomed 2021;34(5):e4257.
156. Alsop DC, Detre JA, Golay X, Gunther M, Hendrikse J, Hernandez-Garcia L, Lu H, MacIntosh BJ, Parkes LM, Smits M, van Osch MJ, Wang DJ, Wong EC, Zaharchuk G. Recommended implementation of arterial spin-labeled perfusion MRI for clinical

applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med 2015;73(1):102-116.
157. Taouli B, Beer AJ, Chenevert T, Collins D, Lehman C, Matos C, Padhani AR, Rosenkrantz AB, Shukla-Dave A, Sigmund E, Tanenbaum L, Thoeny H, Thomassin-Naggara I, Barbieri S, Corcuera-Solano I, Orton M, Partridge SC, Koh DM. Diffusion-weighted imaging outside the brain: Consensus statement from an ISMRM-sponsored workshop. J Magn Reson Imaging 2016;44(3):521-540.
158. Ljimani A, Caroli A, Laustsen C, Francis S, Mendichovszky IA, Bane O, Nery F, Sharma K, Pohlmann A, Dekkers IA, Vallee JP, Derlin K, Notohamiprodjo M, Lim RP, Palmucci S, Serai SD, Periquito J, Wang ZJ, Froeling M, Thoeny HC, Prasad P, Schneider M, Niendorf T, Pullens P, Sourbron S, Sigmund EE. Correction to: Consensus-based technical recommendations for clinical translation of renal diffusion-weighted MRI. MAGMA 2020;33(1):197-198.
159. Ljimani A, Caroli A, Laustsen C, Francis S, Mendichovszky IA, Bane O, Nery F, Sharma K, Pohlmann A, Dekkers IA, Vallee JP, Derlin K, Notohamiprodjo M, Lim RP, Palmucci S, Serai SD, Periquito J, Wang ZJ, Froeling M, Thoeny HC, Prasad P, Schneider M, Niendorf T, Pullens P, Sourbron S, Sigmund EE. Consensus-based technical recommendations for clinical translation of renal diffusion-weighted MRI. MAGMA 2020;33(1):177-195.
160. Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, Martincich L, Partridge SC, Patterson A, Pinker K, Thibault F, Camps-Herrero J, Le Bihan D, group EiBD-WIw. 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):1436-1450.
161. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, Huang W, Noworolski SM, Young RJ, Shiroishi MS, Kim H, Coolens C, Laue H, Chung C, Rosen M, Boss M, Jackson EF. 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(7):e101-e121.
162. Dyverfeldt P, Bissell M, Barker AJ, Bolger AF, Carlhall CJ, Ebbers T, Francios CJ, Frydrychowicz A, Geiger J, Giese D, Hope MD, Kilner PJ, Kozerke S, Myerson S, Neubauer S, Wieben O, Markl M. 4D flow cardiovascular magnetic resonance consensus statement. J Cardiovasc Magn Reson 2015;17:72.
163. Committee QSMCO, Bilgic B, Langkammer C, Marques JP, Meineke J, Milovic C, Schweser F. QSM reconstruction challenge 2.0: Design and report of results. Magn Reson Med 2021;86(3):1241-1255.
164. Karakuzu A, Boudreau M, Duval T, Boshkovski T, Leppert I, Cabana J, Gagnon I, Beliveau P, Pike G, Cohen-Adad J, Stikov N. qMRLab: Quantitative MRI analysis, under one umbrella. Journal of Open Source Software 2020;5(53):2343.
165. Ma D, Gulani V, Seiberlich N, Liu K, Sunshine JL, Duerk JL, Griswold MA. Magnetic resonance fingerprinting. Nature 2013;495(7440):187-192.

Figure Captions:

Figure 1: Example quantitative MR methods illustrated in this manuscript. (Top) Liver proton-density fat fraction (PDFF) mapping, with applications in the evaluation of non-alcoholic fatty liver disease (where liver PDFF is an emerging biomarker for the early diagnosis of NAFLD) and non-alcoholic steatohepatitis (where liver PDFF is emerging as a medical research tool in combination with other noninvasive imaging biomarkers). (Bottom) Cardiac T_1 mapping, with applications in various ischemic and non-ischemic cardiomyopathies.

Figure 2: The major performance metrics for quantitative MR methods are the bias and precision profiles. Bias measures the systematic differences between the measurements and the ground truth. Evaluation of bias requires a highly reliable reference method. Precision measures the tendency of the qMR method to produce different values, when applied repeatedly on the same subject. Evaluation of precision does not require reference values. The term 'profile' refers to the potential variability of performance (bias or precision) across different true values. (SD: standard deviation; CV: coefficient of variation).

Figure 3. Effect of imprecision on the minimum detectable change for a quantitative method. As the within-subject standard deviation (wSD) increases, the minimum detectable change also increases. In this plot, both are given in the units of the measurand. For PDFF, if subjects were always imaged with the same imaging system, we might expect wSD~1%, resulting in a minimum detectable change (95% confidence) of less than 3% absolute PDFF. If subjects were allowed to be imaged with a variety of imaging systems, we might expect wSD closer to 2%, resulting in a minimum detectable change of more than 5% absolute PDFF.

Figure 4: R_2^* decay, if uncorrected, can confound PDFF quantification, leading to bias, as well as poor precision (eg: poor reproducibility across acquisitions with different number of echoes), particularly in patients with elevated liver $R_2^*=1/T_2^*$ ($R_2^*=160$ s^{-1} at 1.5T, corresponding to mild iron overload). As shown though simulation and in vivo, R_2^*-uncorrected signal fitting results are highly dependent on the choice of echo times. In contrast, R_2^*-corrected PDFF quantification has low bias and high reproducibility across choices of echo times. For this illustration, a 12-echo liver CSE acquisition in a patient with high liver fat and iron overload was reprocessed

retrospectively multiple times, using the first n echoes (for n=5,…,12). In each case, both R_2^*-uncorrected and R_2^*-corrected PDFF mapping methods were used.

Figure 5: Technical development and validation is typically an iterative process including technical implementation and refinements, as well as evaluation of bias and/or precision. The later stages of the process are typically more costly, often focused on a specific organ, patient population, and application (context of use), and often include multi-center validation. Once development and validation are completed for one application, extension to other applications is often of interest (e.g., PDFF measurements in skeletal muscle or bone marrow). This extension typically requires additional refinement/validation iterations. Beyond the technical development and validation described in this figure, subsequent clinical qualification (not shown) is needed to establish the relationship between the qMR measurement and specific biological processes or clinical endpoints of interest.

Figure 6: Steps and challenges for the validation, qualification, and dissemination of quantitative MR methods. Technical development and validation as well as clinical qualification are needed in order to establish the performance and clinical utility of qMR methods. Even with successful validation and qualification, substantial challenges (including regulatory and market-based factors) need to be overcome in order to achieve widespread dissemination. Importantly, these three processes can be advanced in parallel as suggested by the horizontal overlap. For example, technical development and validation can be performed at the same time as clinical qualification.

Figure 7: Schematic representation to illustrate that biased measurements can lead to artificially inflated effect size and discrimination between healthy and diseased states, if the confounders of the measurement happen to be sensitive to the pathological alteration. In the example of myocardial T_1 mapping, the most widely used mapping technique MOLLI is known to be confounded by several factors including T_2 time and magnetization transfer (MT) of the tissue. For a specific disease, however, this bias may accentuate the difference between healthy and disease and lead to larger effect sizes compared with unbiased T_1 quantification. However, this gain in effect size comes at the cost of reduced reproducibility as the factors contributing to the bias may vary across acquisitions, systems, and patients.

Supporting Information Table and Figure Captions:

Supporting Information Table S1: Glossary.

Supporting Information Table S2: Individuals and organizations endorsing this manuscript.

Supporting Information Figure S1. Question 1: In what area are the qMR metrics that you study?

Supporting Information Figure S2. Question 2: Please select the confounders you encounter and the frequency with which they appear.

Supporting Information Figure S3. Question 3: What steps do you take to correct confounders, and how often are they implemented in your qMR protocols?

Supporting Information Figure S4. Question 4: Many qMR metrics are promising during the development stage but fail to cross translational gaps. What are the most common hurdles that you face between successful technical development and clinical application? Please evaluate the following in terms of resources required to overcome these hurdles.

Supporting Information Figure S5. Question 5: What tests or metrics do you use to validate a qMR method?

Supporting Information Figure S6. Question 6: What software / packages do you use (or contribute to) for quantitative analysis?

Supporting Information Table S1: Glossary

Qualitative Imaging	The process of obtaining imaging information which is only interpretable in the given imaging context and whose meaning may not be maintained outside the imaging context
Quantitative Imaging	The extraction of quantifiable features from medical images for the assessment of normal or the severity, degree of change, or status of a disease, injury, or chronic condition relative to normal
Measurand	The physical quantity intended to be measured (VIM clause 2.3)*
Performance Metric	A standard measurement of how well a method performs with respect to defined criteria. Common examples include bias and precision.
Bias	A metric describing the systematic measurement error (VIM clause 2.18)*
Bias Profile	A table or figure illustrating the estimates of the bias over the range of true values and/or other relevant characteristics
Precision	The closeness of agreement between measured quantity values obtained by means of replicate measurements of the same MRI system with specified conditions (VIM clause 2.15)*. Repeatability and reproducibility are types of precision (see repeatability, and reproducibility)
Precision Profile	A table or figure illustrating the estimates of the precision (e.g. repeatability or reproducibility) over the range of true values and/or other relevant characteristics
Repeatability	The measurement precision with conditions that remain unchanged between replicate measurements (VIM clause 2.20)*
Repeatability Coefficient	The smallest significant difference between two repeated measurements on a subject when taken under the same conditions
Reproducibility	The measurement precision with conditions that vary between replicate measurements, e.g. different MRI systems (VIM clause 2.25)*
Reproducibility Coefficient	The smallest significant difference between two repeated measurements on a subject when taken under different conditions, e.g. different MRI systems
Ground Truth	The quantity value, consistent with the definition of the quantity (VIM clause 2.11)*. As all measurements are associated with uncertainties this value is considered to be unknowable.
Minimal Detectable Change	The minimum change in a quantity that can be detected with statistical significance.
Reference method	A method that is generally accepted to have suitable performance (low bias and high precision), to be used as a basis for comparison with values of the same kind acquired with the method under investigation (VIM clause 5.18)*

Confounding Factors	Factors that can cause or prevent the outcome of interest, which are not associated with the measurand. This can include the introduction of a bias or the reduction of precision
Technical Validation	The process of demonstrating that the method under investigation meets specified performance criteria. This includes assessment of bias and precision.
Clinical Qualification	Process of demonstrating whether a measurand is linked to a biological process or clinical end point, necessary to demonstrate clinical utility
Biomarker	Characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention, including therapeutic interventions
Clinical Utility	The extent to which a diagnostic method is useful in facilitating beneficial health outcomes (e.g., preventing mortality, enabling early intervention, etc.)
System Imperfections	Deviations of the physical reality of a system from the desired properties. In MRI these may relate to the known and controlled magnetic fields, hardware to transmit, receive, and store RF signals and/or the timing (precision clock).
Harmonization	The process of adjusting of differences and inconsistencies among different systems or methods to make them uniform or mutually compatible
Standardization	The process of developing and implementing standards with the aim of performing comparable methods on different systems
Quality control	Procedures to inspect and verify the quality of the acquired signals, reconstructed images (if any), and measurements. These may include measures of bias or precision obtainable from each acquired dataset.
Quality assurance	Procedures to ensure that the process of measuring a particular biomarker (including imaging equipment, acquisition, and analysis) is adequate

***VIM: International Vocabulary of Metrology (https://jcgm.bipm.org/vim/en/index.html)**

Supporting Information 2:

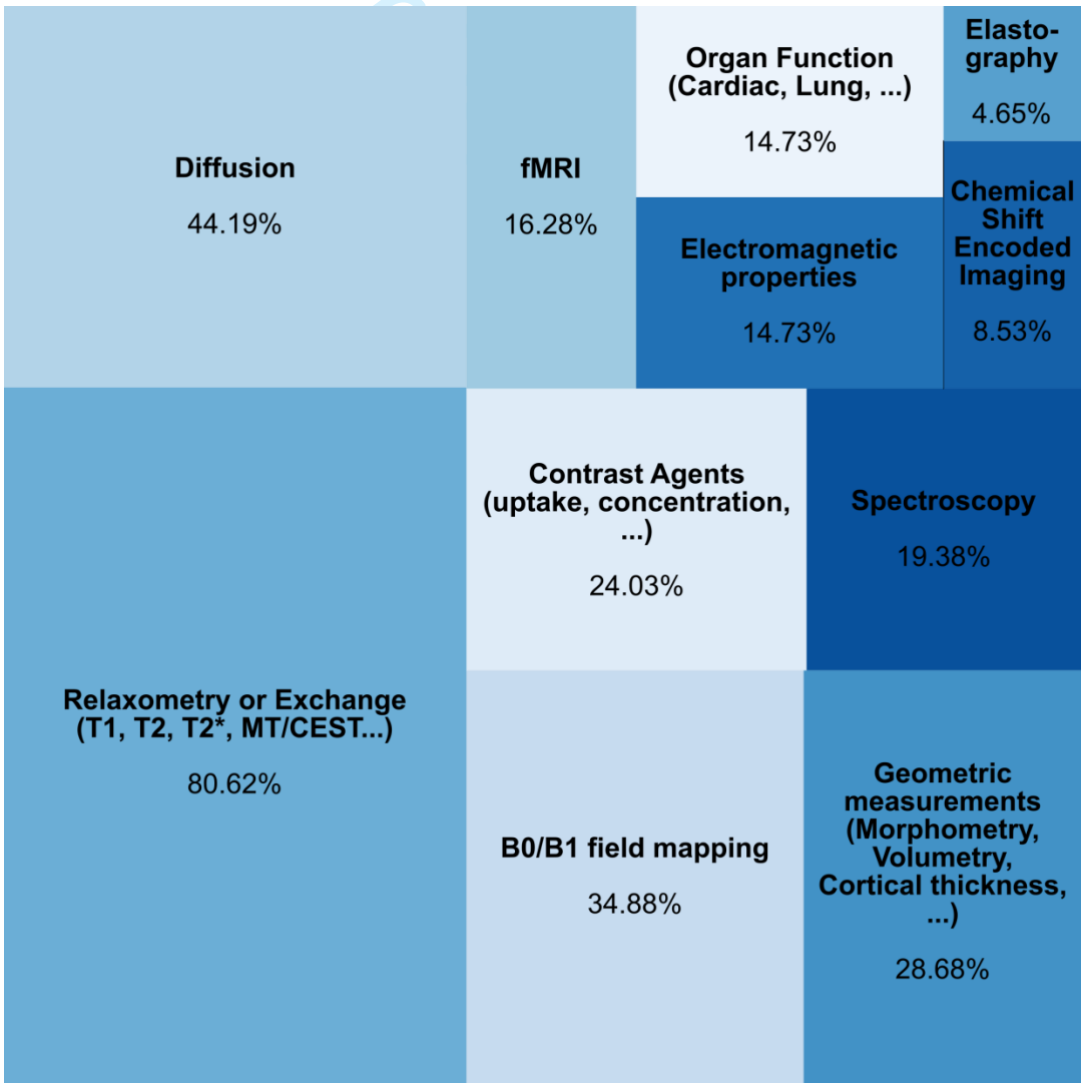
ISMRM qMR study group poll

May 7th 2020

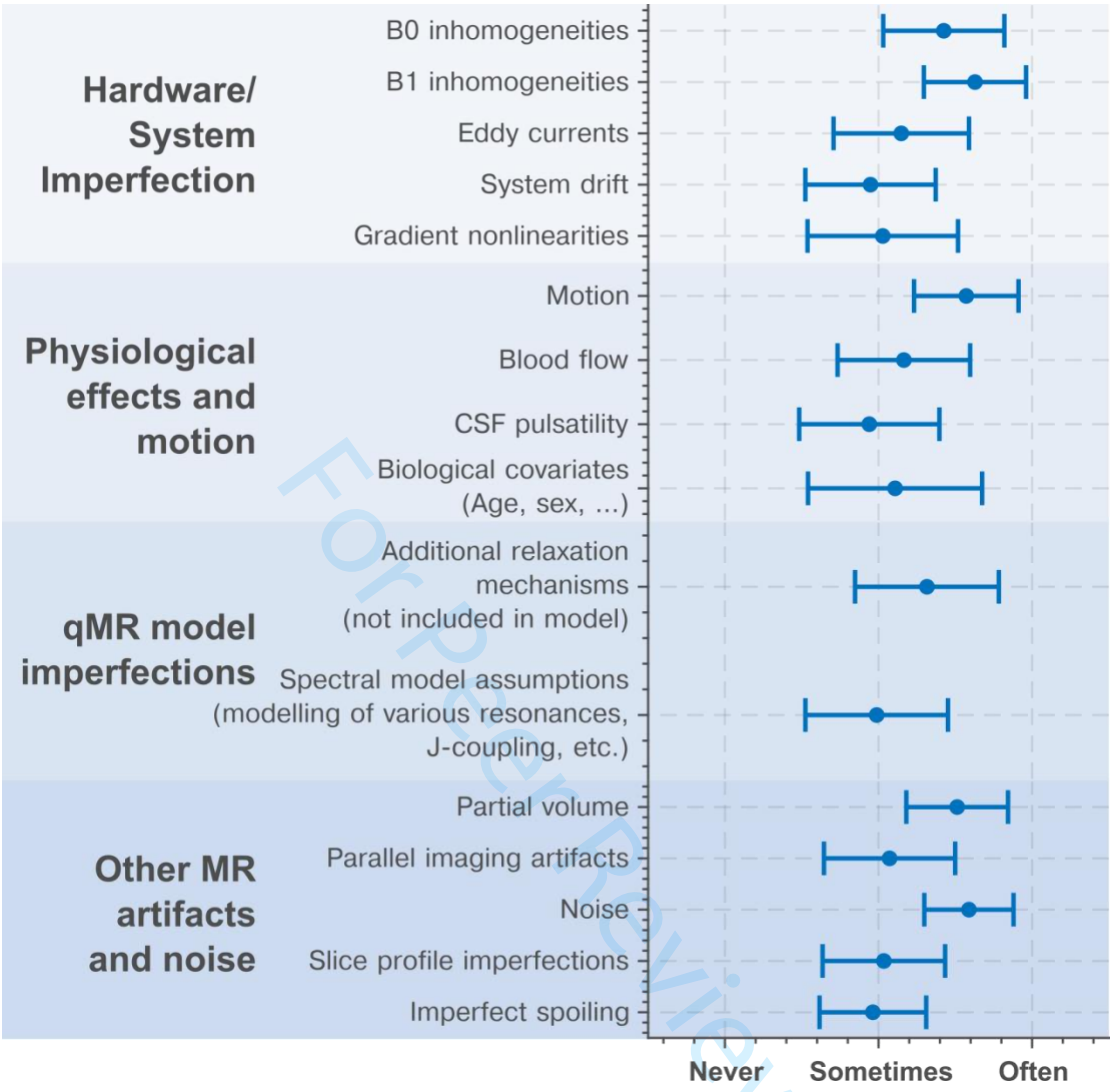
This document summarizes the responses to an informal poll conducted among the ISMRM Quantitative MR Study Group members. The Study Group comprises 1,271 members (as of June 2021), including 908 ISMRM trainee members, 234 ISMRM full members, and 102 SMRT members. Approximately 500 members hold a Ph.D. as their highest degree, 375 a M.Sc., 150 a B.Sc., 40 an M.D., and 23 an M.D./Ph.D. Around 40 countries are represented in the Study Group.

This poll was conducted between February 2020 and May 2020. A total of 129 responses were recorded. In addition to the plots shown below, we have made efforts to include the major themes from the poll responses in the main manuscript text.

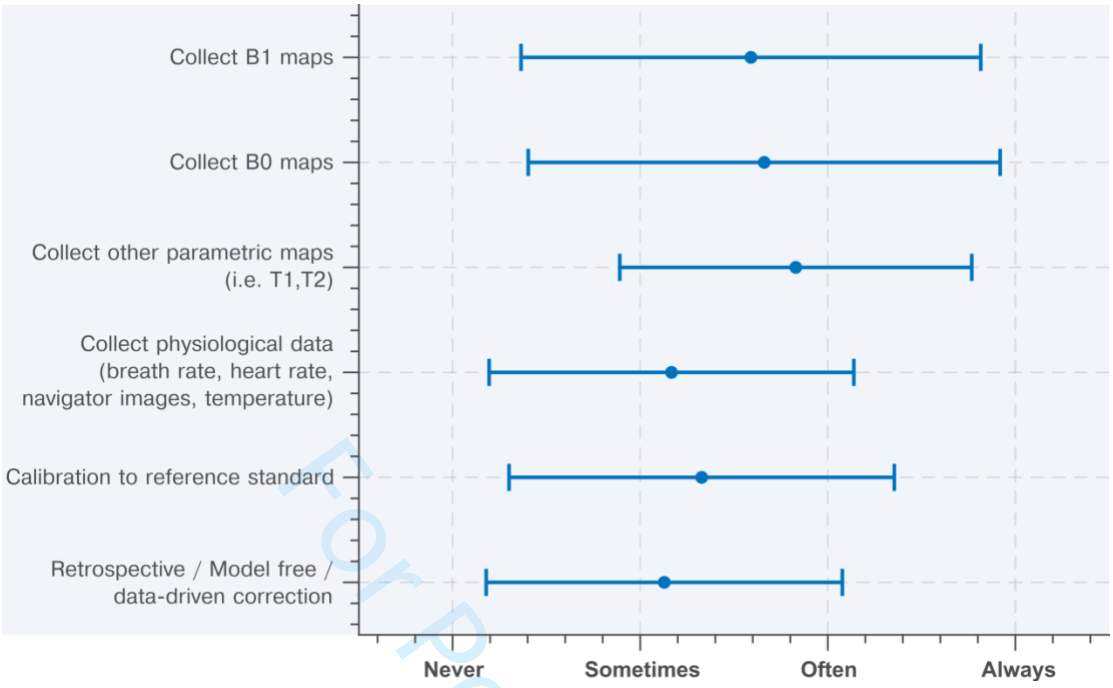
Supporting Information Figure S1. Question 1: *In what area are the qMR metrics that you study?*



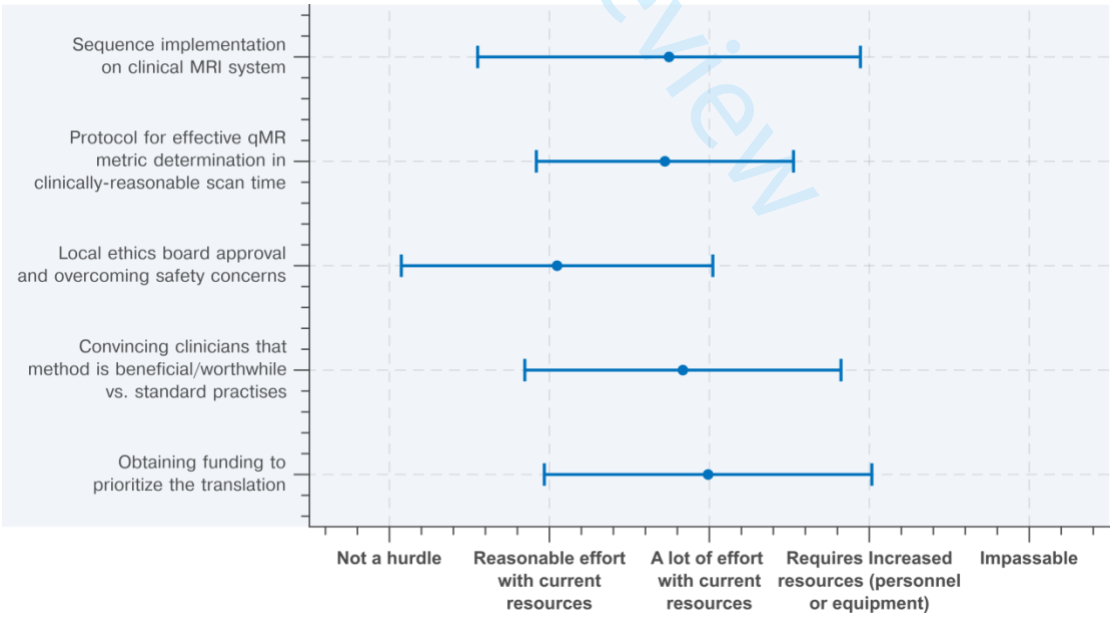
Supporting Information Figure S2. Question 2: *Please select the confounders you encounter and the frequency with which they appear.*



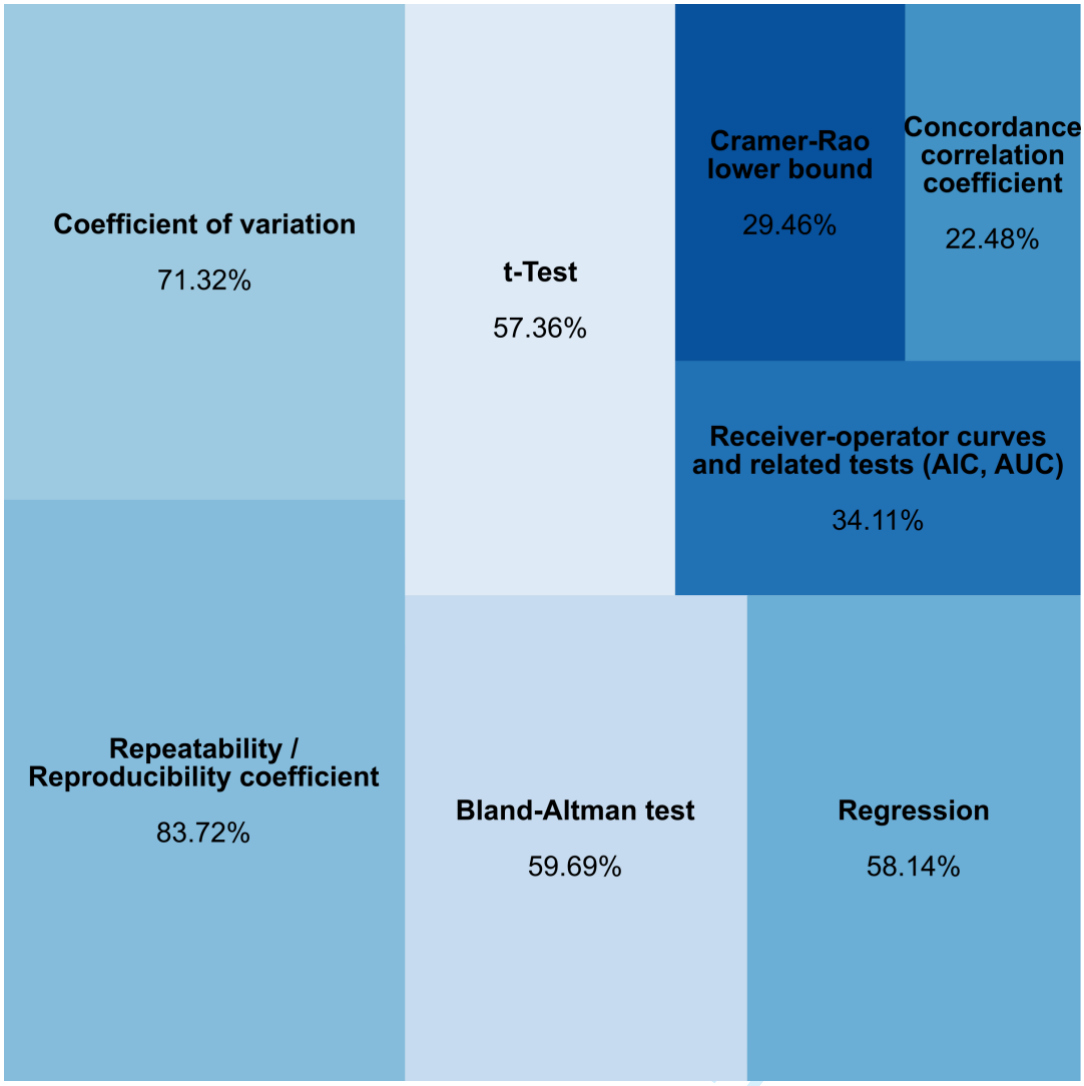
Supporting Information Figure S3. Question 3: *What steps do you take to correct confounders, and how often are they implemented in your qMR protocols?*



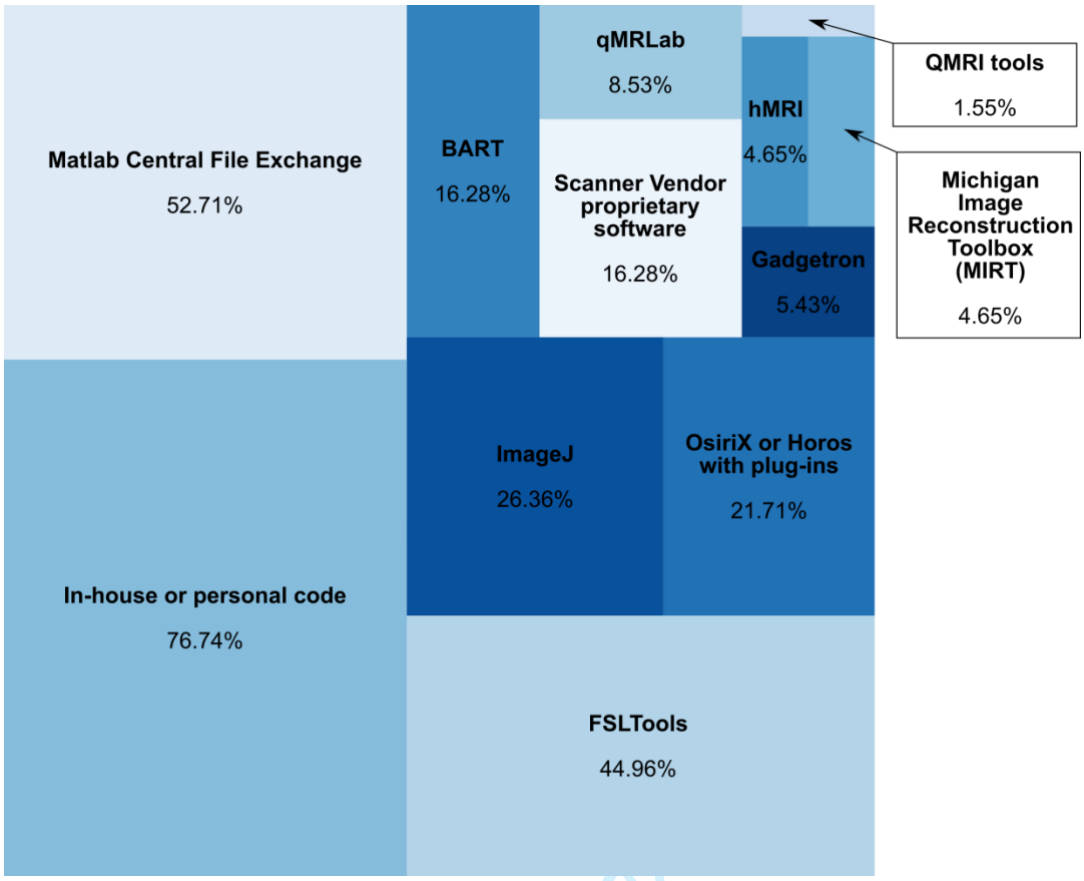
Supporting Information Figure S4. Question 4: *Many qMR metrics are promising during the development stage but fail to cross translational gaps.What are the most common hurdles that you face between successful technical development and clinical application? Please evaluate the following in terms of resources required to overcome:*



Supporting Information Figure S5. Question 5: *What tests or metrics do you use to validate a qMR method?*



Supporting Information Figure S6. Question 6: *What software / packages do you use (or contribute to) for quantitative analysis?*



Supporting Information Table S2: List of Endorsers

To be completed after review

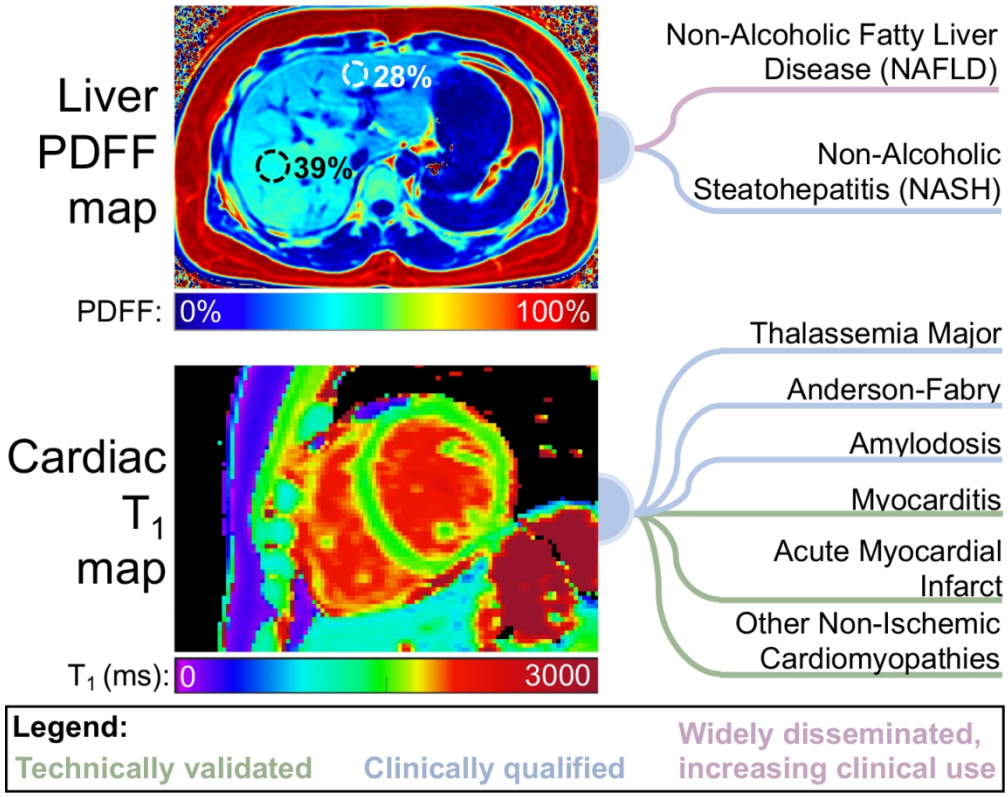


Figure 1: Example quantitative MR methods illustrated in this manuscript. (Top) Liver proton-density fat fraction (PDFF) mapping, with applications in the evaluation of non-alcoholic fatty liver disease (where liver PDFF is an emerging biomarker for the early diagnosis of NAFLD) and non-alcoholic steatohepatitis (where liver PDFF is emerging as a medical research tool in combination with other noninvasive imaging biomarkers). (Bottom) Cardiac T_1 mapping, with applications in various ischemic and non-ischemic cardiomyopathies.

120x95mm (300 x 300 DPI)

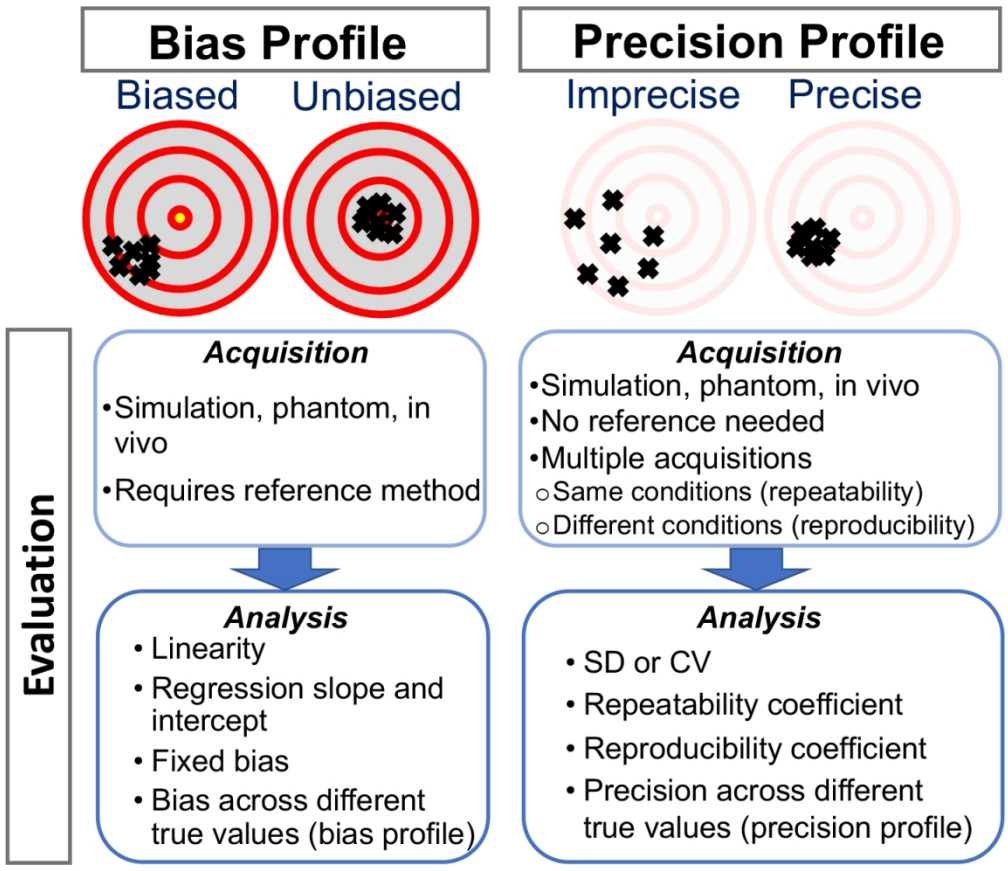


Figure 2: The major performance metrics for quantitative MR methods are the bias and precision profiles. Bias measures the systematic differences between the measurements and the ground truth. Evaluation of bias requires a highly reliable reference method. Precision measures the tendency of the qMR method to produce different values, when applied repeatedly on the same subject. Evaluation of precision does not require reference values. The term 'profile' refers to the potential variability of performance (bias or precision) across different true values. (SD: standard deviation; CV: coefficient of variation).

121x103mm (300 x 300 DPI)

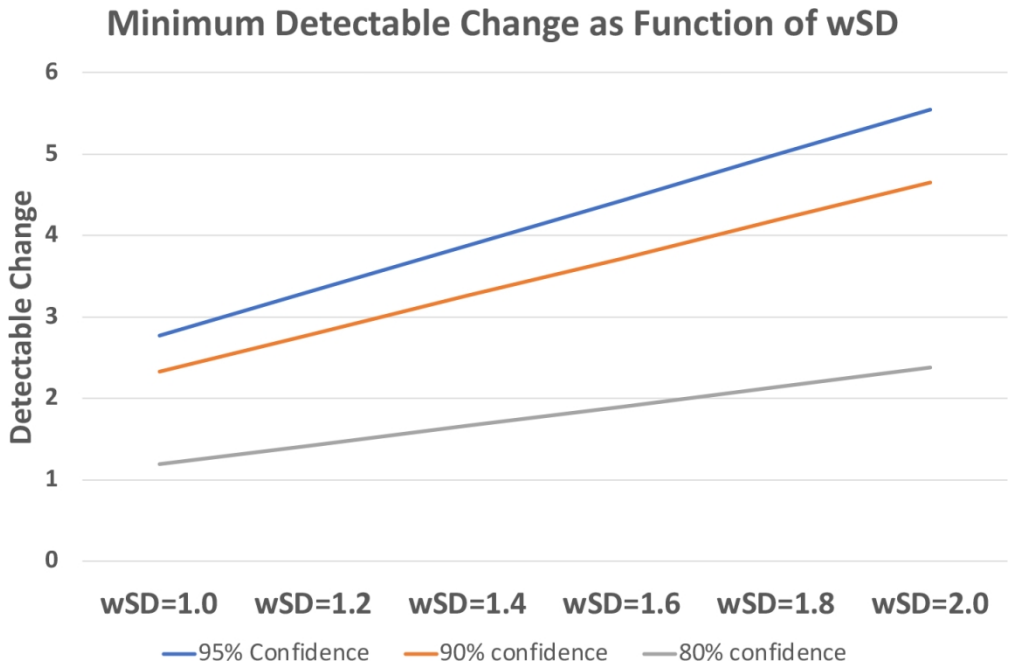


Figure 3. Effect of imprecision on the minimum detectable change for a quantitative method. As the within-subject standard deviation (wSD) increases, the minimum detectable change also increases. In this plot, both are given in the units of the measurand. For PDFF, if subjects were always imaged with the same imaging system, we might expect wSD~1%, resulting in a minimum detectable change (95% confidence) of less than 3% absolute PDFF. If subjects were allowed to be imaged with a variety of imaging systems, we might expect wSD closer to 2%, resulting in a minimum detectable change of more than 5% absolute PDFF.

121x79mm (300 x 300 DPI)

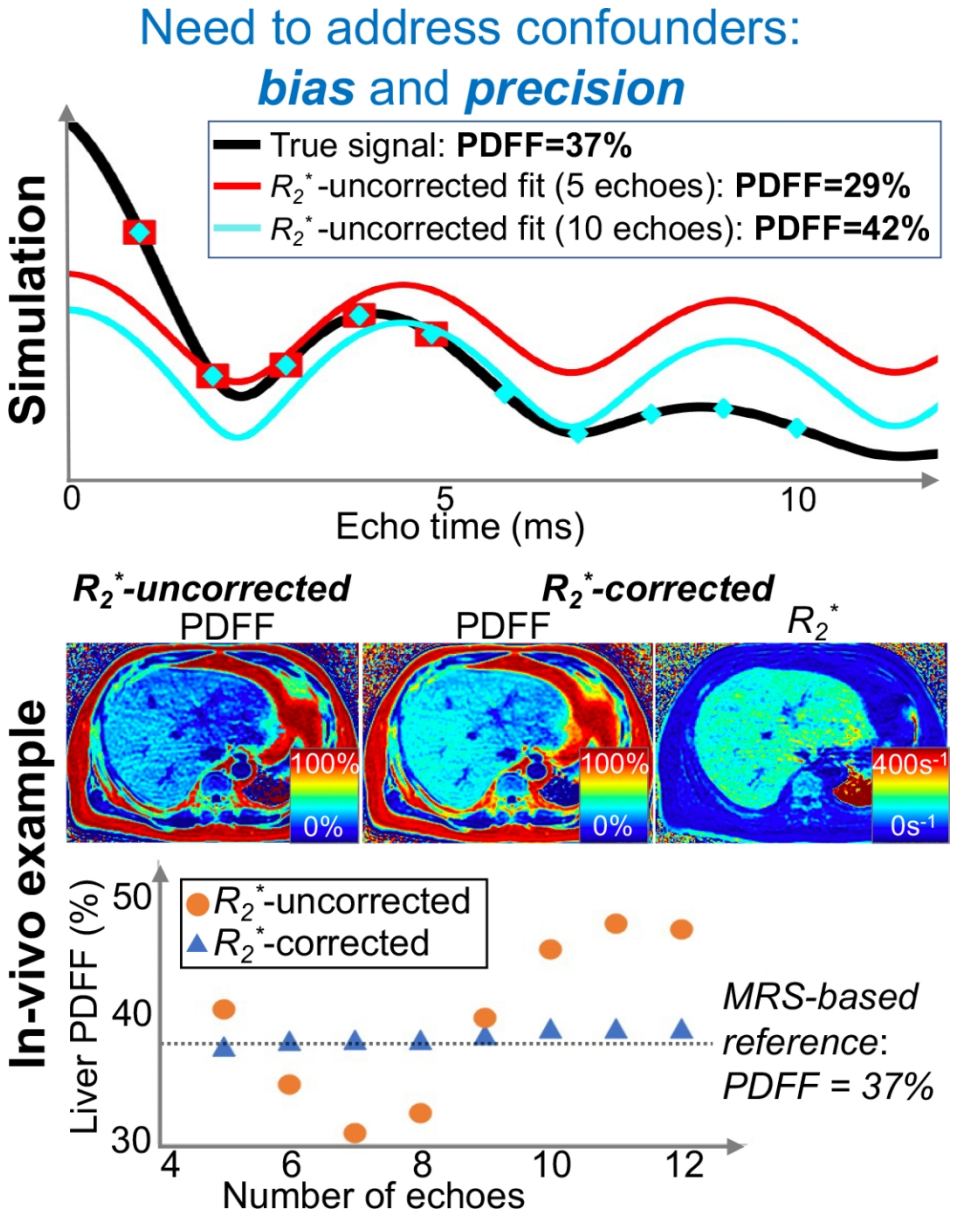


Figure 4: R_2* decay, if uncorrected, can confound PDFF quantification, leading to bias, as well as poor precision (eg: poor reproducibility across acquisitions with different number of echoes), particularly in patients with elevated liver R_2*=1/T_2* (R_2*=160 s^{-1} at 1.5T, corresponding to mild iron overload). As shown though simulation and in vivo, R_2*-uncorrected signal fitting results are highly dependent on the choice of echo times. In contrast, R_2*-corrected PDFF quantification has low bias and high reproducibility across choices of echo times. For this illustration, a 12-echo liver CSE acquisition in a patient with high liver fat and iron overload was reprocessed retrospectively multiple times, using the first n echoes (for n=5,…,12). In each case, both R_2*-uncorrected and R_2*-corrected PDFF mapping methods were used.

95x121mm (300 x 300 DPI)

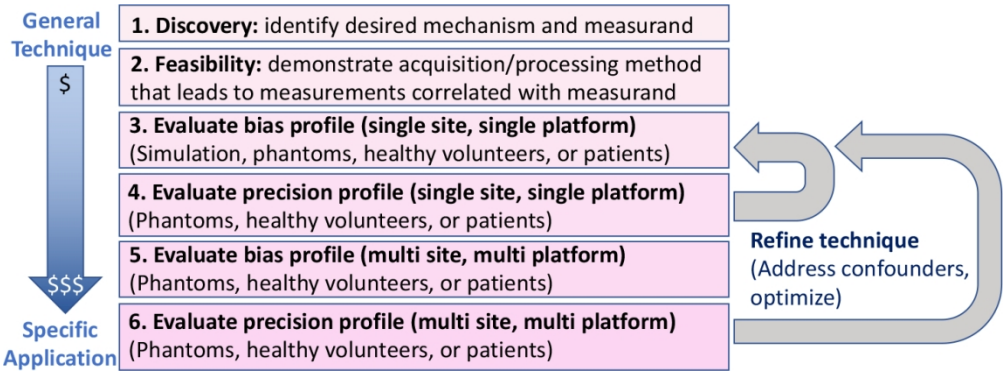


Figure 5: Technical development and validation is typically an iterative process including technical implementation and refinements, as well as evaluation of bias and/or precision. The later stages of the process are typically more costly, often focused on a specific organ, patient population, and application (context of use), and often include multi-center validation. Once development and validation are completed for one application, extension to other applications is often of interest (e.g., PDFF measurements in skeletal muscle or bone marrow). This extension typically requires additional refinement/validation iterations. Beyond the technical development and validation described in this figure, subsequent clinical qualification (not shown) is needed to establish the relationship between the qMR measurement and specific biological processes or clinical endpoints of interest.

122x45mm (300 x 300 DPI)

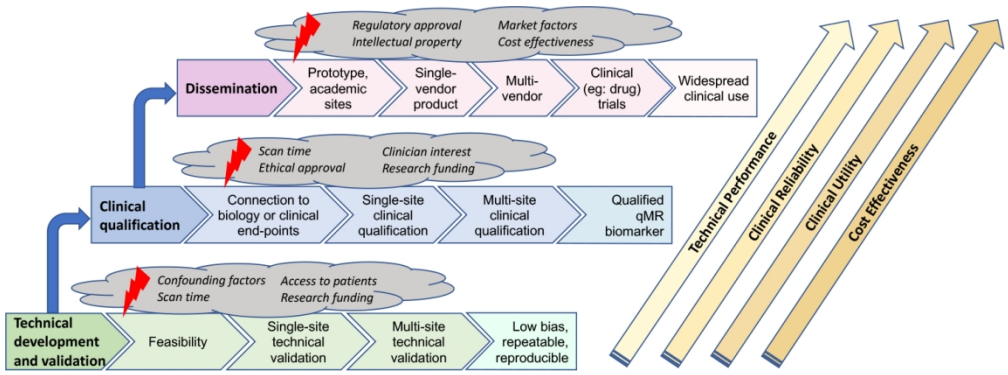


Figure 6: Steps and challenges for the validation, qualification, and dissemination of quantitative MR methods. Technical development and validation as well as clinical qualification are needed in order to establish the performance and clinical utility of qMR methods. Even with successful validation and qualification, substantial challenges (including regulatory and market-based factors) need to be overcome in order to achieve widespread dissemination. Importantly, these three processes can be advanced in parallel as suggested by the horizontal overlap. For example, technical development and validation can be performed at the same time as clinical qualification.

126x46mm (300 x 300 DPI)

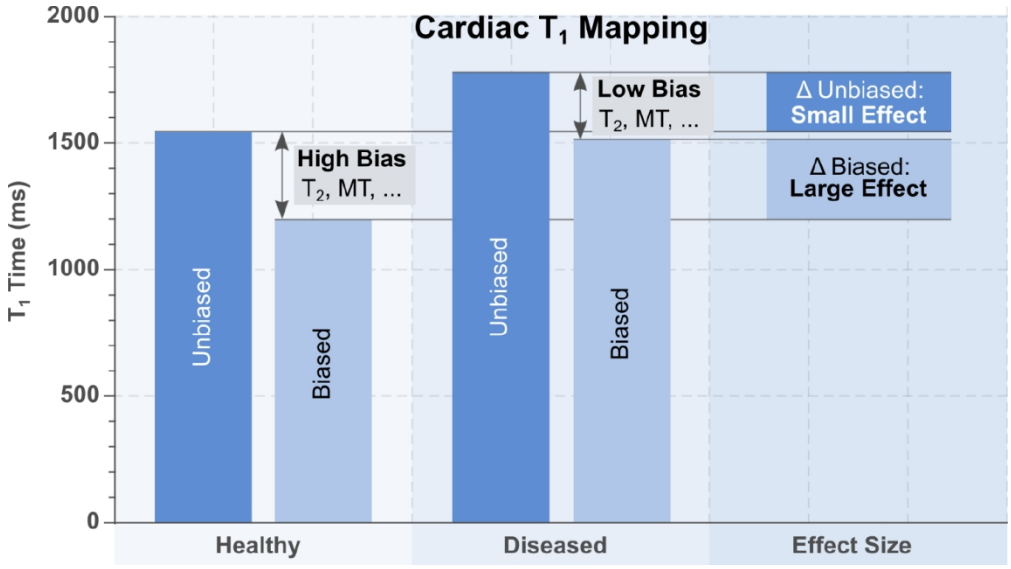


Figure 7: Schematic representation to illustrate that biased measurements can lead to artificially inflated effect size and discrimination between healthy and diseased states, if the confounders of the measurement happen to be sensitive to the pathological alteration. In the example of myocardial T_1 mapping, the most widely used mapping technique MOLLI is known to be confounded by several factors including T_2 time and magnetization transfer (MT) of the tissue. For a specific disease, however, this bias may accentuate the difference between healthy and disease and lead to larger effect sizes compared with unbiased T_1 quantification. However, this gain in effect size comes at the cost of reduced reproducibility as the factors contributing to the bias may vary across acquisitions, systems, and patients.

126x70mm (300 x 300 DPI)