Quantitative imaging through the production chain from idea to application

Salameh, Najat; Weingärtner, Sebastian; Hilbert, Tom; Vilgrain, Valérie; Robson, Matthew D.; Marques, José P.

DOI
10.1007/s10334-023-01131-w

Publication date
2023

Document Version
Final published version

Published in
Magnetic Resonance Materials in Physics, Biology and Medicine

Citation (APA)
Salameh, N., Weingärtner, S., Hilbert, T., Vilgrain, V., Robson, M. D., & Marques, J. P. (2023). Quantitative imaging through the production chain: from idea to application. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *36*(6), 851-855. https://doi.org/10.1007/s10334-023-01131-w

Magnetic Resonance Materials in Physics, Biology and Medicine (2023) 36:851–855
https://doi.org/10.1007/s10334-023-01131-w

COMMENTARY

Quantitative imaging through the production chain: from idea to application

Najat Salameh[1] · Sebastian Weingärtner[2,3] · Tom Hilbert[4,5,6] · Valérie Vilgrain[7] · Matthew D. Robson[8] · José P. Marques[9]

Received: 20 October 2023 / Revised: 25 October 2023 / Accepted: 27 October 2023 / Published online: 11 November 2023

Magnetic resonance imaging (MRI) has revolutionised radiology, offering a wealth of contrast information on the structure, function and metabolism of organs. Yet, clinical MR images used in routine protocols rely on shades of grey where hypointensities and hyperintensities are visually identified by radiologists. Quantitative MR imaging (qMRI) is a promising alternative providing reproducible measurements that can be compared across patients and timepoints. With quantitative imaging, we are referring not only to relaxometry-type measurements, where water spin magnetic properties are probed, but also any physical quantities that can be derived from MRI signal by means of biophysical models (with diffusion, perfusion, fat fraction, flow and quantitative susceptibility mapping being some of the most notable examples). Since quantitative imaging typically requires longer acquisition times than qualitative imaging and because its full clinical value and meaning are not yet fully understood in all pathologies, its clinical adoption is often met with some resistance. Modern MR technology including high-performance hardware and computing power has allowed scientists in academia and the industry to effectively accelerate the acquisition and reconstruction processes, making quantitative imaging increasingly compatible with medical diagnostic workflows.

Whilst every scientific conference revolving around MR research systematically showcases a large variety of new methods to improve and validate quantitation, these are not broadly adopted by stakeholders such as MR vendors, radiologists, digital medical technology companies or CROs (Contract Research Organisations). Introducing quantitative measures into the diagnostic pipeline is not trivial and necessitates that all stakeholders converge in answering what, why and how a new quantitative marker should be measured and used to diagnose or evaluate a disease state (or its progression).

Having this in mind, four of these key stakeholders were invited at the 2023 annual meeting of the ESMRMB to discuss our responsibilities in taking quantitative imaging from the drawing board to clinical practise: Academia; MR Industry; Clinical researchers; Companies running clinical trials. In this commentary article (as in the dedicated educational session), the four panellists were asked to reflect independently on disclosing and reporting a new quantitative method (academia), the process a new MRI method has to undergo to become a product (MR industry), as well as what are the most important requirements for a quantitative method to be used in either clinical validation studies, routine diagnosis (radiology) or clinical trials (CROs or pharmaceutical companies). Those highly interconnected aspects are summarised in Fig. 1 and already illustrate how much the different stakeholders, usually composed of multi-disciplinary teams,

✉ José P. Marques
Jose.marques@donders.ru.nl

1 Center for Adaptable MRI Technology (AMT Center), Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK

2 Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands

3 HollandPTC, Delft, The Netherlands

4 Advanced Clinical Imaging Technology, Siemens Healthineers International AG, Lausanne, Switzerland

5 Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

6 LTS5, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

7 Department of Radiology, Hôpital Beaujon, APHP.Nord, Université Paris Cité, Paris, France

8 Perspectum Ltd, Oxford, UK

9 Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands

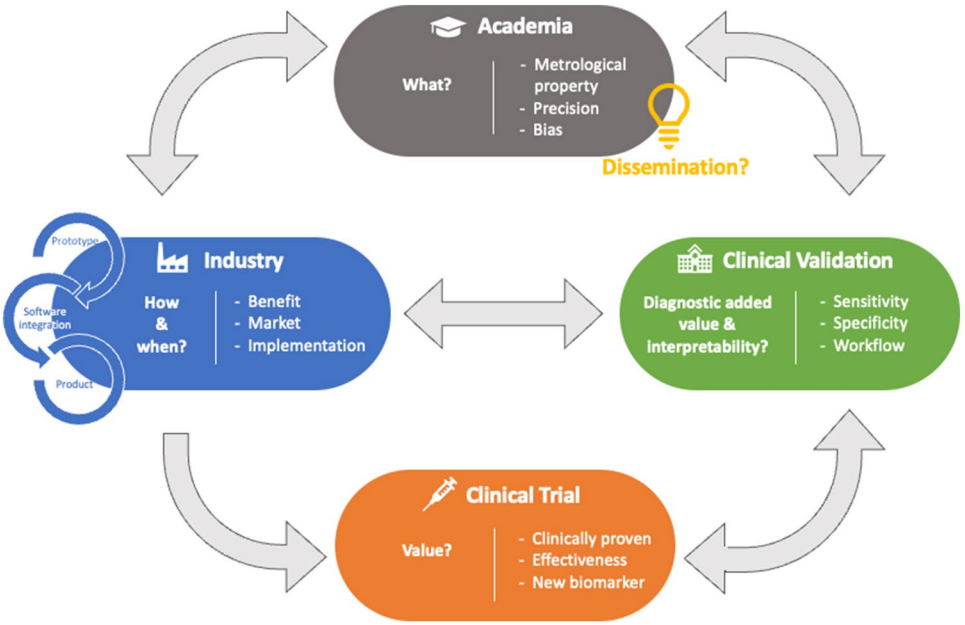

Fig. 1 Illustration describing the links and iterative developments between the various stakeholders

have to work in synergy to reach one common goal: bringing state-of-the-art diagnostics to the patients. Finally, we will report on the many points raised during the Roundtable discussion, both by our speakers and participating audience.

Academia: what is a quantitative method?

To set the development of a new quantitative method up for success, thorough testing, and validation is the most integral component. This process begins when evaluating if the new development is indeed a quantitative measurement technique. Looking at the international metrological vocabulary, a verbose definition tells us that a measurement is quantitative if it retains its meaning outside a given context and measures a well-defined property. When measuring this property, the so-called measurand, errors are made. In metrology, the error is decomposed into two components: (1) the random measurement error, or precision, which averages out with many adequate repetitions, and (2) the systematic measurement error, or bias, which describes the deviation that remains in the absence of the random measurement error. Precision can be assessed by test–retest measurements and, thus, is often best performed in vivo. Bias on the other hand requires an estimate of the ground truth. To this end, imaging researchers usually find reference techniques that measure the same measurand but with much lower bias and much better precision. Reference techniques are often only available in phantoms, which is thus the most common choice for assessing bias. To create confidence in the metrological characterisation of a method, it is most convincingly provided to the community in an open-science manner, along with the underlying data and methods, to reproduce the results.

Bias and precision metrics are integral to understanding the intrinsic limitations of our measurement techniques [1]. Even though there is merit in assessing these abstract metrics, a measurement will drive clinical use only if it can detect or stage a certain condition. However, these concepts are closely related. Precision is directly proportional to the minimal detectable change and the cohort size required for demonstrating an effect, at any given statistical certainty. The relationship between bias and clinical metric on the other hand is less straightforward. Small biases can be well tolerated, particularly if they are small compared to the random error or if thorough harmonisation is ensured and appropriate reference ranges are used. However, bias usually stems from uncorrected confounders and is often subject to change when the underlying confounders vary. Thus, large bias is generally the number one contributor to poor reproducibility, which can often only be identified late in the testing and development cycle, e.g. when bringing the technique to multi-centre/multi-vendor evaluation.

To ascertain these clinical metrics, thorough clinical testing is indispensable for translation. This step usually requires multi-disciplinary collaboration. This will involve identifying the right partners, and successfully sharing the methods with those partners. In this process, the clinical reproducibility will be put to the test and potential clinical utility will be identified. Ideally, this evaluation forms the starting point of the long journey of getting the quantitative biomarker used in clinical practise.

Ultimately, comprehensive characterisation requires a range of investigations, starting from metrological testing

to obtain bias and precision, to clinical testing to obtain an understanding of how this affects sensitivity and specificity. Only understanding those limits of a method allows adequate interpretation of its meaning and puts it to good clinical use.

Clinical validation: what is the added diagnostic value and interpretability of quantitation in the clinic?

Fat quantitation is a good example illustrating the use of quantitative MRI in clinical studies. Metabolic dysfunction-associated steatotic liver disease (MASLD) affects over 25% of the global population and is a common cause of chronic liver disease [2]. Hepatic steatosis, defined as the abnormal accumulation of lipids in the cytoplasm of hepatocytes, is one of the main features of MASLD and is detected by both histology and non-invasive imaging.

The classical approach to measure fat within the liver is the MRI-proton density fat fraction (MRI-PDFF), which is the ratio of the mobile proton density from triglycerides and the total mobile proton density from triglycerides and water. It has been known for more than a decade that MRI-PDFF correlates well with MR spectroscopy and is more sensitive than the histology-determined steatosis grade in quantifying increases or decreases in the liver fat content [3]. Since then, it has been shown that significant MRI-PDFF changes are correlated to clinical outcomes in MASDL patients. In particular, increased liver fat content equal or superior to 15% is associated with increased odds of fibrosis progression at an early stage of fibrosis whilst a 30% or more MRI-PDFF decline relative to baseline is associated with a histological response such as an improvement in the activity score and fibrosis regression [4].

Stopping here, one could think that MR fat fraction is a perfect, successful example of quantitation used in the clinic. Nevertheless, despite these encouraging findings, the association between liver fat content, consequently MRI-PDFF and liver-related outcomes and mortality remain complex. First, liver fibrosis is the most important prognostic factor. Second, liver fat content decreases in the setting of MASLD patients with cirrhosis. Therefore, taking MASLD patients with all fibrosis stages, steatosis is not associated with prognosis. This point was well illustrated in a study with patients having extensive fibrosis or cirrhosis which showed that steatosis grade < 33% was associated with higher incidence of liver-related events and death [5].

So where does MRI-PDFF stand in MASLD patients today? Certainly not as the diagnostic method of choice and the reasons are twofold. First, because there are numerous confounding factors and steatosis alone is not the only pathologic finding of interest but is associated with inflammation, fibrosis and carcinogenesis—thus, it lacks specificity. Second, because qualitative and/or quantitative ultrasound is able to, at a lower cost and being more accessible, stratify patients with no or minimal steatosis vs. the others. Nevertheless, as MRI-PDFF changes capture histological changes, MRI-PDFF is commonly used as an end-point to evaluate treatment effects in early phase clinical trials. Yet, MRI-PDFF is not used to monitor MASLD patients in clinical practise. To reach that goal, an MRI technique would certainly benefit: (i) from being based on a multiparametric approach to create a combined biomarker that would increase its discriminatory power towards confounding factors and increase the specificity; from being more cost competitive to increase its accessibility.

Industry: how and when is a new quantitative method ready to become a product?

The path from an idea to a product can be challenging for MRI imaging methods. Additional hurdles can be found for methods that provide a quantitative value, as is the case for quantitative mapping (e.g. T1/T2 mapping) or morphometry (e.g. brain structure volumes). There is no specific time point when a method is evaluated as suitable for the product, but various criteria should be considered throughout the lifetime of the method. The lifetime of a method can be crudely divided into three phases after the idea has been conceived:

1. Prototype
2. Software integration
3. Product

The most important and longest phase is the prototype phase which typically starts with a publication that describes the method and performs a first technical validation in terms of accuracy and precision. With the distribution of the prototype, it can be iteratively improved in terms of user-friendliness and robustness. As examples, in Siemens, these prototypes are named Work-in-Progress [WIP] and Customer to Customer packages [C2P], whilst in the Philips architecture, these are software "patches". At the prototyping phase, the method's clinical feasibility can be evaluated in clinical research, e.g. how robust is the method towards motion in patients or are the acquisition and reconstruction times acceptable for clinical workflow? In the same setting, additional validation studies, published as peer-reviewed articles, are crucial to stay on the path to an end-product. These collaborative publications are important since only few internal validations are typically performed before a decision for productisation is made. Furthermore, studies investigating different organs, patient demographics and

testing repeatability/reproducibility are decisive for regulatory clearance at the end of productisation.

In the software integration phase, the prototype will have to be re-implemented and tested in the product code for future release on the main scanner software. Yet, entering this phase is not straightforward and multiple aspects must be investigated. The one that prevails is the overall amount of time required for software integration? In that regard, key questions must be answered: For example, are there safety relevant features of the sequence that must be changed? Is additional hardware required? Does an online image reconstruction exist? In practise, how many work hours will it take to integrate the new feature and how much will it increase the complexity of the existing product code? This whole endeavour will then be weighed against the potential benefits of releasing the method as a product. To that end, a market analysis is performed by interviewing key opinion leaders, running a literature search and surveys. For qMRI, this can be a challenging task since a widespread clinical use is yet to come, effectively reducing the market for such products.

Before a method can finally become a product, it needs to be cleared by regulatory bodies in the respective countries where it is intended to be used. Typically, specific requirements apply for quantitative imaging methods and involve additional validation studies and extensive documentation as it is, for example, described by the Food and Drug Administration [6].

Clinical trial: what is a clinical trial looking for and what value could bring quantitative methods?

A clinical trial is looking to explore the impact of a pharmaceutical company's drug on a patient. This might be to de-risk their project, or evaluate the effectiveness of the agent so that they can present the results to a regulatory body. Quantitative methods should be appealing because they directly support statistical comparison. Introducing a new method to a clinical trial is feasible, only if the following three aspects are carefully considered.

Requirements for a new method to be used in a clinical trial

For a method to be used in a trial, the drug company has to be able to trust the results. In practise, this means that there needs to be evidence that the measurement from the method is robust, reliable, relevant and trustworthy. The best measurements would be approved biomarkers already used clinically, or in pharmaceutical trials, some of these are accepted as surrogate endpoints [7]. Peer-reviewed publications for the method and methods that measure physical quantities would both be favoured, but completely new metrics are rarely welcome. Whilst the regulators can be consulted in advance when exploring a new measurement or approach, they are typically extremely conservative as they have the safety of the patient as their priority. In summary, drug trialists do not want uncertainty from the measurement methods, they need methods that they can take to regulators and that the regulators will not question later.

Considerations on multisite data, standardisation of acquisitions across vendors

The methods need to be deployable and some level of technical homogenisation is valuable as patients' images, obtained from different sites and scanner brands, will be pooled. Large trials can involve hundreds of centres and all need to be equipped with the same methods. The latter have, hence, to be available on all the standard clinical MRI systems, independently of the manufacturer, and at both 1.5 T and 3 T. In addition, these sites, which are often in multiple countries, need to be supported for training, data transfer, quality control, software upgrades, etc. If a new method needs non-standard product acquisition strategies (be it in the form of special "patches" or C2P/WIPs in the case of Philips or Siemens pre-development), this would translate in additional paperwork and the incurred delays would become overwhelming and dissuasive.

Importance of specificity and sensitivity

The trial design will dictate the needs of the biomarker, and MRI-based markers are favoured when possible as MRI is non-invasive. To be considered in a clinical trial, a new marker is deemed efficient if it can deliver quick and reliable answers regarding: 1—patient's stratification, and/or 2—the disease progression or treatment response. To that end, and to answer first the question of inclusion vs. exclusion of a patient (perhaps to enrich the amount of diseased patients in a study), then sensitivity or specificity would be important metrics. To monitor treatment response in trials including large numbers of patients (100 s or 1000 s), then repeatability is a key metric. In the typical case, as the MRI-PDFF example illustrated above, that aims at evaluating the changes in individual patient's liver PDFF over time, group changes will be assessed. In this case, the bias from one timepoint to the next is more of a concern than the random errors in the measurement. Thus, quantitative methods do not need to be perfect (simultaneously having high sensitivity, specificity and repeatability) but they do need to be well characterised.

Conclusion

In summary, all panellists could agree on the need to develop quantitative tools that would not only help to screen and triage patients, but also stage the disease and assess the body response to treatment. To achieve these goals, they recommend fostering communication channels between stakeholders to ensure that each party understands the needs of each other in the production chain. Although challenging to address, some points have been highlighted that we, as a community, should embrace to improve both clinical relevance and adoption of qMRI. In particular:

-Academics would ease translation if their methods remain simple, easy to use and are clinically relevant;

-Clinicians could facilitate new developments and method deployment by taking advantage of their role as opinion leaders;

-MR manufacturers would stimulate transfer to clinical use if methods are standardised and transparent;

-CROs would facilitate method validation if they could act as facilitators between method developers and pharmaceutical companies.

Some of those points are heavily intertwined, to the point that they could appear unsolvable, placing all stakeholders in a "chicken and egg" dilemma. Although all panellists as well as the audience appreciated the difficulty of breaking out of such situations, they all acknowledged a path to success, a recipe that includes three main ingredients: dialogue, frictions and resilience. Dialogue will help understanding the needs and constraints of each partner; frictions will help cross-discipline, scientific exchanges, and iterations to tune the new methods to those needs; resilience: well, we do not think this point needs further explanation, science is a fantastic, yet never-ending journey!

Declarations

Conflict of interest Tom Hilbert is an employee of Siemens Healthineers, Matthew D. Robson is an employee and CTO of Perspectum Ltd, and the remaining co-authors have no conflicts of interest.

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

References

1. Weingärtner S et al (2022) Development, validation, qualification, and dissemination of quantitative MR methods: overview and recommendations by the ISMRM quantitative MR study group. Magn Reson Med 87(3):1184–1206. https://doi.org/10.1002/mrm.29084
2. Younossi Z et al (2018) Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 15(1):11–20. https://doi.org/10.1038/nrgastro.2017.109
3. Noureddin M et al (2013) Clinical and histological determinants of nonalcoholic steatohepatitis and advanced fibrosis in elderly patients. Hepatol Baltim Md 58(5):1644–1654. https://doi.org/10.1002/hep.26465
4. Tamaki N, Ajmera V, Loomba R (2022) Non-invasive methods for imaging hepatic steatosis and their clinical importance in NAFLD. Nat Rev Endocrinol 18(1):55–66. https://doi.org/10.1038/s41574-021-00584-0
5. Vilar-Gomez E et al (2018) Fibrosis severity as a determinant of cause-specific mortality in patients with advanced nonalcoholic fatty liver disease: a multi-national cohort study. Gastroenterology 155(2):443-457.e17. https://doi.org/10.1053/j.gastro.2018.04.034
6. C. for D. and R. Health (2023) Technical performance assessment of quantitative imaging in radiological device premarket submissions. Accessed 23 Oct 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/technical-performance-assessment-quantitative-imaging-radiological-device-premarket-submissions
7. C. for D. E. and Research (2022) Table of surrogate endpoints that were the basis of drug approval or licensure. FDA. Accessed 23 Oct 2023. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure