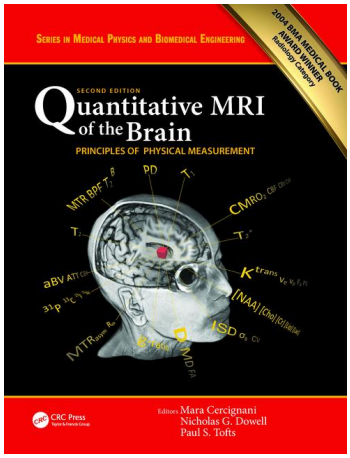
Author's preprint.

The published version of this chapter is in the book:
Quantitative MRI of the Brain: Principles of Physical Measurement, Second edition
Editors: Mara Cercignani, Nicholas G. Dowell, Paul S. Tofts
CRC press: ISBN: ISBN 9781138032859

September 18th 2018



Chapter 1: concepts: measurement in MRI[1]

Paul S Tofts
Brighton and Sussex Medical School, Brighton, United Kingdom.

Table of Contents will go here

[1] Reviewed by Mara Cercignani

1. Introduction

1.1. Measurement science and MRI come together.

Measurement science has been around a long time; MRI[2] has been around for about 35 years. This book is about the blending of the two paradigms.

We have come to expect to be able to measure certain quantities with great accuracy, precision and convenience. Instruments for mass, length and time are all conveniently available, and we expect the results to be reproducible when measured again, and also to be comparable with measurements made by others in other locations. In the human body we expect to measure some parameters (height, weight, blood pressure) ourselves, recognising that some of these parameters may have genuine biological variation with time. More invasive measurements (e.g. blood alcohol level, or blood sugar level) are also expected to have a well-defined normal range, and to be reproducible. In physics, chemistry, electrical engineering and manufacturing industry there is a strong tradition of measurement, international agreements on standards, and training courses for laboratory practitioners. International standards of length, mass and time have been in existence for many years. Secondary standards are produced, which can be traced back to the primary standards. National and international bodies provide coordination.

As individual scientists we may have a passionate desire to use our talents for the benefit of mankind, preferring to devote our energy to finding better ways of helping our fellow humans to be healthy than to improving weapons for their destruction. In this context, developing measurement techniques in MRI constitutes a perfect application of traditional scientific skills to a modern problem.

MRI is now widespread, and accepted as the imaging method of choice for the brain (and for many body studies). It is generally used in a qualitative way, with the images being reported non-numerically by radiologists. Many MRI machines now have independent workstations, connected to the scanner and the database of MR images, which enable and encourage simple quantitative analysis of the images in their numerical (i.e. digital) form. However the data collection procedure often prevents proper quantification being carried out; variation in machine parameters such as transmitter output gain, flip angle value (and its spatial variation), receiver gain, and image scaling may all be acceptable for qualitative analysis, but cause irreversible confusion in images to be quantified. Researchers may be unaware of good practice in quantification, and collect or analyse data in an unsuitable way, even though the MRI machine is capable of more.

The process of quantifying, or measuring parameters in the brain necessarily takes more time and effort than a straightforward qualitative study. More MRI scanner time is needed, and considerable physics development effort and computing resources may be needed to set up the procedure. In addition, analysis can be very time consuming, and support of the procedure is required to measure and maintain its reliability over time. Procedures have to be found which are insensitive to operator procedure (whether in the data collection or image analysis) and to scanner imperfections (such as radiofrequency nonuniformity from a particular head coil), which provide good coverage of the brain in a reasonable time, and which are stable over study times which may extend to decades.

The benefits of quantification are that fundamental research into biological changes in disease, and their response to potential treatments, can proceed in a more satisfactory way. Problems of bias, reproducibility and interpretation are substantially reduced**.** MRI can move from a process of picture-taking, where reports are made on the basis of unusually bright, dark, small or large objects, to a process of measurement, in the tradition of

[2] Magnetic Resonance Imaging is a term invented by US radiologists to describe Nuclear Magnetic Resonance (NMR) imaging. The 'nuclear' part was removed from the name NMR to prevent the public being alarmed. NMR spectroscopy (chapter 12) was originally concerned with identifying chemical compounds, and there was no spatial information contained in the data. It developed separately from imaging, on different machines, and is often referred to as MRS. Modern MRS is carried out largely on MRI machines, and uses the imaging gradients to localise the spectra to particular parts of the body. For these reasons, MRI is now considered to include spectroscopy. MR is a more correct term, and refers to MRI and MRS together.

scientific instrumentation, where a whole range of quantities can be tested to see whether they lie in a normal range, and whether they have changed from the time of a previous examination.

In this book, the intention is to demonstrate the merging these two traditions, or paradigms, of measurement and of MRI, to form the field of quantitative MRI, or qMRI[3]. The MRI measurement process is analysed, often in great detail. Limits to accuracy and precision are identified, as far as possible, with the intention of identifying methods that are reliable and yet practical in a clinical MRI scanning environment. The biological meaning of the many MR parameters that are available is explored, and clinical examples are given where MR parameters are altered in disease. Often these changes have been observed qualitatively, and they serve to encourage us to improve the measurement techniques, in order that more subtle effect of disease can be seen, earlier than currently possible, and in tissue that is currently thought to be normal as judged by conventional MRI. The ideal is to obtain push-button (turnkey) techniques for each of the many MRI parameters in this book, such that a MRI radiographer (technologist) can measure each of these parameters reliably and reproducibly, with a minimum of human training or intervention, in the same way that we can currently step onto a weighing machine and obtain a digital readout of our mass. In the case of qMRI the output would be considerably richer, perhaps showing images of abnormal areas (computed from large databases of normal image datasets), changes from a previous MRI exam, possible interpretations (diagnoses), and an indication of certainty for each piece of information. The advances in the pre-scan and the spectroscopy MR procedures, which used to be very time-consuming and operator-dependent, and are now available as fully-automated options, show how this might be possible.

Thus MRI has been undergoing a *paradigm shift*[4] in how it is viewed and used. In the past it was used for forming qualitative images (the 'happy-snappy MRI camera', taking pictures); in the future it may be increasingly used a scientific instrument to make measurements of clinically relevant quantities. The dichotomy can be seen in the MRI literature; radiological descriptions often speak of signal hyperintensity in a sequence with a particular weighting, whilst studies using physical measurements often report localised concentration values, normal ranges, age and gender effects, and reproducibility. As measurement becomes more precise, and analysis enables clinically relevant information to be extracted from a myriad of information, in will become possible, in principle, to make measurements on an individual patient to characterise the state of their tissue, guiding the choice of treatment and measuring its effect. The issues involved in bringing qMRI into the radiological clinic were well summarised in an editorial in the American Journal of Neuroradiology (McGowan, 2001) see box 1.

As part of this ongoing paradigm shift, our view of what MRI can tell us is changing. When it started, information was largely anatomical (*anatomical MRI*), in the sense that relatively large structures would be observed. Changes in their geometric characteristics (usually size), compared to normal subjects, or to a scan carried out in previous weeks or months, would be noted. Quantitative examples would be volume and atrophy. *Functional MRI* (fMRI) claimed the complementary ground, studying short-term changes in tissue arising from carrying out particular (neural) functions. *Micro-structural MRI* occupies a third role, as shown in this book. Many MR parameters (such as diffusion, magnetisation transfer, spectroscopy) show structural changes in tissue arising from damage caused by disease at a microscopic level. To observe these changes directly would require imaging resolution of the order of 1-100μm[5], since they generally involve a variety of biological changes at the cellular level.

[3] The website www.qmri.org can be used for updates

[4] Thomas Kuhn, in The Structure of Scientific Revolutions, first introduced the idea of paradigm shifts. An example would be the move from a classical physics to a quantum physics view of the world. A paradigm is a pattern or model, a way of viewing the world or part of it, a point of view, a mindset.

[5] 1 *micron* (μm) is 10^{-3} mm or 10^{-6} m.

There are a growing number of quantitative MR applications that represent evolutionary change in the use of MR imaging. These applications also include magnetization transfer techniques, absolute T1 and T2 measurements, functional imaging, and a number of spectroscopic techniques.

A significant challenge in the clinical employment of quantitative methods is that underlying physical mechanisms may not yet be fully understood in the context of what can be measured with the MR imaging experiment. For example, one can associate the presence of abnormalities in quantitative measures with the presence of disease, but causality may not be established.

Thus, results are sometimes limited to empirical findings of correlation with some other measure or observable process. Still such results are potentially of great value by providing means of noninvasive disease characterization and, thus, insight into the natural history of disease.

Another substantial benefit is derived from the use of validated methods to study the efficacy of novel therapeutic agents. Coupled with results of other studies, including investigations in animal models in which correlation may be observed between the results of an invasive or destructive test and the results of noninvasive MR imaging, human studies such as the present investigation serve to connect clinical observation with imaging findings.

It is statistically advantageous to follow up preliminary studies that use ‘‘many’’ measures with targeted studies that have the power to accept or reject the hypothesis that certain measures are significantly correlated.

Investigators differing from the authors of the original study may do this only when precise and comprehensive data regarding the study methods are provided. However, even when the authors make a good-faith effort to disclose every nuance of the experimental method, it still may be difficult to control for differences in MR hardware and software. This is in part because modern MR system design objectives are focused on obtaining excellent-quality clinical images for conventional, subjective interpretation.

Adapted from McGowan 2001

Box 1 Bringing qMRI into the radiological clinic

These can be observed by pathologists in post-mortem tissue, using optical or electron microscopy and special staining techniques (histopathology). This resolution is well below the spatial resolution of MRI (which is about 1 mm on clinical scanners). However changes at the microscopic level (e.g. in cellular structure) give changes in the MR parameters (e.g. in water diffusion) which can be observed at coarser spatial resolution (of about 1 mm). Thus structural changes of sizes well below those that would be called anatomical can be detected. In addition, the concentrations of chemical compounds (metabolites) in cells, and their changes, can be measured with spectroscopy. The physiological permeability of the endothelial membrane around blood vessels can be measured using dynamic imaging of Gd-contrast agent.

These changes may occur both in a so called ‘*lesion*’, which is tissue seen at post-mortem and in conventional MRI to be visibly different from the surrounding tissue, and in the ‘*normal-appearing’ tissue*, which appears normal in conventional MRI. Lesions are usually described as ‘*focal*’, meaning that the change is localised to a relatively small area (a few mm or cm), with a distinct boundary; thus its differing brightness in an image distinguishes it from the surrounding tissue (considered normal). In contrast, a diffuse change may extend over more area, has no distinct boundary, and is harder to detect by simple visual observation of the image. Diffuse changes are often well characterised by quantification, since it is the absolute value of quantities within the area that is measured, without reference to surrounding tissue, or the need for a distinct boundary.

1.2. Limits to progress

It may appear that qMRI research proceeds under its own impetus. However the current state and rate of progress in developing reliable qMRI methodology are determined by several factors: MRI manufacturers, research institutions, pharmaceutical companies, computer and electronics technology and publicly funded research councils.

MRI machine manufacturers (vendors) will take on some of the measurement procedures over time, incorporating them into their research and development programs, and then offering them as turnkey (push-button) products[6]. The speed of this process is driven by demand from clinical purchasers, by whether competing manufacturers offer such facilities, and by whether public medical funding bodies such as the US Food and Drugs Administration (FDA) is likely to approve re-imbursement of the cost of such procedures from medical insurance policies. The existence of a large and growing installed base of high-quality, reliable and ever improving MRI machines, primarily designed for routine clinical use, largely in environments where they can be run as parts of profitable businesses, has enabled and encouraged the development on these machines of qMRI techniques, although they are still of interest to only a (growing) minority of users[7]. As MRI machines evolve, the qMRI techniques usually have to be re-implemented.

Research institutions have particular structural strengths and weaknesses. qMRI needs input from chemists, computer scientists, neurologists, physicists, radiologists and statisticians. There may be good career support for those applying methods to study clinical problems, but none for those basic scientists inventing and developing the methods. There may be a clash of paradigms or traditions, between those who have been educated in a hierarchical environment where asking questions is considered to be irrelevant or subversive, and those who consider asking questions to be an absolute basic necessity of undertaking modern high-quality scientific research. The availability of talented researchers in turn depends on how much value is placed on science in society, schools and universities, and whether appropriate postgraduate training opportunities exist. The International Society for Magnetic Resonance in Medicine (ISMRM)[8] is a powerful force bringing together researchers from different institutions who are working on similar methodologies, through both its journals and its scientific meetings,

The demand from pharmaceutical companies and neurologists for qMRI measurements to be used in drug trials is large and likely to increase (Miller, 2002) (Filippi and Grossman, 2002) (Filippi *et al.*, 2002; McFarland *et al.*, 2002) (Sormani *et al.*, 2011; Mallik *et al.*, 2014) . The traditional double-blind placebo-controlled phase III trial involves many patients (typically 100-1000) being studied for several years in order to obtain enough statistical power to determine whether a drug is effective. The large sample size is needed to deal with the variability of disease in the absence of treatment, and the imperfect treatment effect (which may vary according to patient subgroup). Such trials typically cost several US $100 million. qMRI can potentially shorten the procedure, by identifying treatment failures early on in the testing process, on a smaller sample. If there is no observed biological effect from the treatment, it may be considered unlikely that the drug is working (this will depend on the particular way the drug has been postulated to act). For example, if a potential treatment for multiple sclerosis (MS) showed no effect on all the MR measures that are known to be abnormal in MS, it would probably be dropped in favour of other drugs. With new biotechnology and gene-based treatments being developed, the number of candidate drugs for evaluation will increase by a large factor, and traditional trials will become too expensive and slow to evaluate all of them. Thus direct in-vivo qMRI observation of treatment effect could become increasingly valued.

The rapid increase in power and availability of computing technology has also been key in enabling data acquisition and image analysis techniques to be realised. Numerically designed magnets, coils and radiofrequency

[6] These are often sold as extras

[7] The 'killer app' can sometimes galvanise action around making a qMRI parameter available. This is when an application is found that has a clear clinical importance (e.g. MD in stroke).

[8] www.ismrm.org

pulses, digital receivers and rapid image registration and analysis have all changed the way that MRI is carried out.

The resources available from pharmaceutical companies to drive the process of developing and supporting reliable qMRI measures may exceed those available from traditional publicly funded research sources. Traditional research council sources have been willing to support the application of qMRI methods to study particular diseases, but often unwilling to support the development of new quantitative methods, sometimes claiming that MRI manufacturers should be doing this.

1.3. Using this book

The field of quantitative MRI should include the following four key areas: Basic concepts of measurement, how to measure each MR parameter (to include both acquisition and analysis), and the biological significance of each parameter (with input from post mortem and possibly animal studies). For each MR parameter, the following aspects are important:

i) the biological significance of the MR parameter

ii) how it can be measured accurately and slowly

iii) how it can be measured practically and quickly

iv) examples of clinical applications

v) what can go wrong in the measurement procedures

vi) QA approaches (controls and phantoms)

vii) normal values for tissue

viii) reproducibility performance that can be achieved

ix) multi-centre studies

x) future developments.

The editors did their best to get authors to consider all these ten aspects in their chapters.

The book is intended to be a repository of qMRI methodology, of particular use to PhD students; hopefully the methodology will not need to be reinvented by each generation of researchers. The first edition of this book (Tofts, 2003) contains some information not present in this edition which may be worth consulting; the chapter authors in this edition have changed and necessarily give a different perspective.

In chapters 2 and 3 the issues in measurement that occur repeatedly throughout the book, as each MR parameter is considered, are examined in more detail. These are grouped into the processes of data collection, data analysis and quality assurance, all of which crucially affect how well MR quantities can be measured. Units are usually given in SI (System International), and conventions used in this book for physical units and symbols (e.g. TR,TE,T_1,T_2) are those recommended in the style guide for the journal Magnetic Resonance in Medicine, published for the International Society for Magnetic Resonance in Medicine[9]. Most of the focus is on techniques which can be implemented on standard clinical MRI scanners; some techniques (e.g. ^{31}P spectroscopy or ^{23}Na imaging) need non-standard hardware as an add-on.

This book is intended for researchers who already have a basic knowledge of how Magnetic Resonance Imaging works, and some knowledge of the brain, including the major diseases (cancer, epilepsy, stroke, multiple sclerosis and dementia). Newcomers can find many appropriate books (table 1), and also helpful websites such as ISMRM.

2. History of measurement

2.1. Early measurement

Early quantitative techniques focussed around the desire to measure distance, mass, monetary value and time. An awareness of these can give us perspective in our own endeavours to quantify!

Developed in about 3000 BC in ancient Egypt, the *cubit* was a ubiquitous standard of linear measurement, equal to 524mm. It was based on the length of the arm from the elbow to the extended fingertips and was standardised by a royal master cubit of black granite, against which all cubit sticks used in Egypt were to be measured at regular intervals[10]. The precision of the thousands of cubit sticks used in building the great Pyramid of Giza is thought to have been very high, given that the sides of the pyramid are identical to within 0.05%.

Early *astronomers* developed remarkably precise measurement methods (as demonstrated at Stonehenge); their ability to guide navigation and predict eclipses brought them fame. In the 16th century, precise

[9] see http://www.ismrm.org/journals.htm

[10] much of the historical material in this chapter comes from the Encyclopaedia Britannica.

calculations of planetary orbits by Copernicus, Kepler and Galileo challenged the intellectual dominance of the Catholic Church, bringing an end to the idea that all heavenly bodies rotate around the earth.

title	authors	date published	no. of pages	description
Magnetic Resonance Imaging: physical principles and sequence design	RW Brown, YC Norman Cheng, EM Haacke, MR Thompson, R Venkatesan	2014	976	thorough exposition of MRI principles; 2nd edition
MRI from picture to proton	DW McRobbie, EA Moore, MJ Graves	2017	400	written by experienced physicists; new edition
Quantitative MRI in cancer	TE Yankeelov, DR Pickens, RR Price	2011	338	Multi-author book by radiologists and physicists; a 'sister book' to this one
Quantitative MRI of the spinal cord	J Cohen-Adad, C Wheeler-Kingshott	2014	330	Multi-author, 'sister book' to this one
Diffusion MRI: theory, methods and applications	DK Jones	2011	784	Covers much of quantitative brain MRI from a physics point of view
Handbook of MRI pulse sequences	M Bernstein, K King, X Zhou	2004	1040	A pulse programmer's friend
MRI in Practice	C Westbrook, CK Roth, J Talbot	2011	456	Established book giving radiographer's viewpoint
Barr's The Human Nervous System: an anatomical viewpoint	JA Kiernan, R Rajakumar	2013	448	includes complete description of the brain

Table 1. Recommended books for background reading in MRI and neuroanatomy
NB Statistics books are in chapter 2 (table 2).

In 1581 the word *quantitative*[11] was first used, meaning 'involving the measurement of quantity or amount'. Quantity means 'size, magnitude or dimension' in Middle English. In 1847 'quantitative analysis' was used, meaning 'chemical analysis designed to determine the amounts or proportions of the components of a substance'. In 1878 *quantify* was used to mean 'to determine the quantity of, to measure', and hence *quantification* is 'the operation of quantifying'. In 1927 '*quantitate*' was used to mean 'to measure or estimate the quantity of, especially to measure or determine precisely'. However Webster's calls this term a 'back-formation[12]', which is probably as derogatory

[11] see Webster's dictionary and the Oxford English Dictionary

[12] a *back-formation* is a word formed by subtraction of a real or supposed affix from an already existing longer word. Thus from *quantitation* was created *quantitate*.

as a dictionary compiler can be, and this term is not used in this book, nor is it in the Oxford English Dictionary.

Francis Bacon (1561-1626) had a great influence on generations of British scientists who followed him (Gribbin, 2003). He stressed collecting as much data as possible, then setting out to explain the observations, instead of dreaming up an idea and then looking for facts to support it. Science must be built on the foundation provided by the facts. What would he say about the modern 'hypothesis-drive' research? In 1662, the Royal Society of London for the Promotion of Natural Knowledge received its charter from King James II, as one of the first (and best known) scientific societies.

2.2. The Longitude problem – John Harrison

In the 18th century the problem of navigation around the globe was severe. Although latitude (distance from the equator) could be measured accurately, using the elevation of the sun above the horizon at noon (the time of maximum altitude), longitude (the easterly or westerly distance around the globe, now measured from Greenwich, London, UK) could not be (Sobel, 2005). Samuel Pepys, commenting on the pathetic state of navigation, had written of "the confusion all these people are in, how to make good their reckonings, even each man's with itself", recognising the distinction between intra- and inter-observer variation. Newton wrote of the sources of error involved in trying to measure time at sea "One [method for determining longitude] is by a Watch to keep time exactly. But, by reason of motion of the Ship, the Variation of Heat and Cold, Wet and Dry, and the Difference of Gravity in different Latitudes, such a watch hath not yet been made".

As a result many lives were lost at sea, through shipwreck and failure of supplies, and navigation was such a sensitive issue that sailors were forbidden to carry out their own calculations, for fear that they would show up errors in those of their superior officers. The growth of vastly profitable world trade was held back. In this context, the Longitude Act of 1714 was passed in the British Parliament, offering a reward of £10,000[13] to anyone who could devise a method of measuring longitude accurately.

The challenge of solving the 'longitude problem', as it came to be known, was taken up by an English clockmaker, John Harrison, who lived near the port of Hull, and had heard the stories of souls go to their death, and the reward offered. He built four clocks altogether. The first kept good time on land (better than 1 second per month) and in small trips out to sea. The Longitude Board could give incentive awards to help impoverished inventors bring promising ideas to fruition. He succeeded in getting a full trial at sea with the navy, on a voyage to Lisbon in 1736; his clock showed unexpected error at sea, being susceptible to an artefact caused by accelerations in the motion at sea. His own perfectionism, and obstinacy all round, delayed matters, and the next trial, taking his fourth clock to the West Indies, did not take place for another 25 years. The Longitude Board was dominated by eminent astronomers and others from the naval establishment, and repeatedly refused to give Harrison his payment, requiring that the chronometer should first be taken from prototype into mass productiont. The Board realised that replicate voyages and clocks were needed to establish the reproducibility, without which the accuracy could not be guaranteed. A single measurement could not establish the maximum error. His son William took up his case, and the Royal Society offered him Fellowship. It was only intervention by King George III, and the passing of a second act by Parliament, that gave him his recognition, at the age of 80, 46 years after he had built his first sea clock.

This story, of finding a scientific solution to a human problem, has all the elements of the struggles that modern scientists may have to develop a technique that they believe will save lives, and many parallels can be seen. Harrison's clocks are preserved in the old Royal Observatory at Greenwich.

2.3. Scientific societies

The Lunar Society of Birmingham (England) was a group of forward thinking scientists who met between 1766 and 1791. They met on the day of the full moon (so that

[13] the sum was graded according to the accuracy that could be achieved.

travel would be easier), and flourished independently of the Royal Society (in London). Birmingham was the location of much inventive scientific activity stimulated by the industrial revolution. Both of Charles Darwin's grandfathers (Josiah Wedgewood, the pottery manufacturer and Erasmus Darwin, the naturalist) were members, as were Matthew Boulton (the manufacturer), Joseph Priestly (who discovered oxygen) and James Watt (who invented the steam engine). The Industrial revolution in Britain and the rest of Europe gave commercial impetus to the invention of a variety of measuring instruments to be used in the manufacturing process. Lord Kelvin, delivering a lecture on electrical units of measurement in 1883, expressed the desire of his time to quantify:

> *"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be."*

although he might have added a caveat about the danger of numbers giving a pseudo-scientific respectability to some studies.

2.4. Units of measurement

In the newly formed United States of America, it was found impossible to reform the archaic system of weights and measures inherited from the British, in spite of the Napoleonic metric system that had recently been adopted in France. The Office of Weights and Standards became the National Bureau of Standards, then the National Institute of Standards and Technology (NIST). In 1960 the 11th general Conference of Weights and Measures, meeting in Paris, established the International System of Units, based on the metre, kilogram, second, ampere, degree Kelvin, and candela. These units are often called the *SI units*, after the French expression *Système Internationale*, and are preferred in the scientific community[14]. The kilogram is represented by a cylinder of platinum-iridium alloy kept at the International Bureau of Weights and Measures in France[15], with a duplicate in the USA; the other units are defined with respect to natural standards (e.g. the metre is defined by the wavelength of a particular visible atomic spectral line). National centres such as the US NIST and the UK National Physical Laboratory (NPL) are now centres of expertise in measurement science.

2.5. Mathematical physics

In parallel with the development of physical instruments had been the discovery of mathematical techniques. Ancient Babylonians, Egyptians, Greeks, Indians (Harappans) and Chinese all had mathematics, originally used for computing areas and volumes of regular objects, and also used for handling monetary currency. In the 6th century BC, Pythagoras established the link between the musical note of a string, and its length. This bridge between the world of physical experience and that of numerical relationships has been called the birth of mathematical physics, where numbers explain the origin of physical forms and qualities. Newton's differential calculus and Fourier's transform are essential tools used by our current MRI scanners. Early digital computers, most famously used to decipher the Enigma code used by submarines during the Second World War, developed to the stage we take for granted today.

2.6. Scientific medicine

In medicine the concepts of the new scientific methods, including quantification, were applied. William Harvey (1578-1657) was a physician and scientist who studied the blood circulation extensively, and was the first to measure the cardiac volume and estimate the total blood volume in the human body. In 1833 William Beaumont, a US army surgeon, published a series of studies[16] on a soldier who

[14] the engineering community in the USA still uses units based on the British Imperial system (although these are not used in the UK any more). Incompatibility between Imperial and Metric units was blamed for a space vehicle failure in the late 1990's.

[15] the BIPM, Bureau International des Poids et Mesures http://www.bipm.org/en/about-us/

[16] from The Man with a Lid on his Stomach, in the Faber Book of Science, edited by John Carey.

had been wounded in the stomach and then developed a flap that could be opened. Beaumont could watch food in the stomach, and extract gastric juice. Nowadays we have more convenient ways of making in-vivo studies.

In the late 1970's scientists started connecting medical imaging hardware to computers that look extremely basic by modern standards, motivated by the desire to manipulate and interrogate the images. Sophisticated medical imaging instruments were produced, in nuclear medicine, ultrasound, X-ray Computed Tomography, and Nuclear Magnetic Resonance.

In about 1978 the annual meetings on Information Processing in Medical Imaging (IPMI) started taking place. In 1989 it was argued that attention to good data collection was at least as important as sophisticated image processing (box 2). The notion that good quantification required attention to both *data collection* and *image analysis* techniques was born, and this complementarity can be seen in the structure of this book. Experience has shown that advances are often made by groups who have access to both data collection (so that the acquisition technique can be optimised for the job in hand) and to advanced analysis techniques (to obtain the most from

The history of image processing in nuclear medicine shows that collection of good quality image data is at least as important as access to image processing techniques. Even now one could argue that real improvements in the usefulness of image data come from instrumental improvements rather than from more sophisticated ways of image processing. However in the case of large datasets that are already of good quality, the problem is then one of data presentation and reduction, rather than correcting images to compensate for errors in data collection.

With this philosophy we have initially concentrated on collecting good quality data, that are sensitive to the clinical question being studied. For example T_2 weighted images of the brain can show Multiple Sclerosis lesions, and one could develop sophisticated algorithms for measuring lesion volume to assess disease and therapies; however the images show oedema and scar tissue, which are secondary to the disease process. Primary visualisation of the disease is shown by the newer technique of GD-DTPA scanning, and therefore we have developed this data collection technique in preference. A second example is the use of expensive classification techniques on image data clearly showing gross nonuniformity which can be removed relatively simply.

Having taken care of the instrumental aspects and obtained good quality data, the processing requirements may become less expensive, and mostly consist of PACS, 3D display, calculation of functional images, and segmentation algorithms. Where sophisticated forms of information processing are required, to make full use of them they must be integrated into a programme that includes aspects of data collection such as sequence design, quality control of instrumental parameters, validation of the quantitative results, and good experimental design. In summary, we believe that data must be *appropriate,* and of *good quality,* before undertaking any processing.

Box 2: A plea for 'good quality data collection'. PACS is Picture Archival and Computing System, and refers to computer based systems to store, display and interrogate large quantities of medical images. By 'functional images' was meant parametric maps of any kind (e.g. permeability). From (Tofts *et al.*, 1991a; Tofts *et al.*, 1991b)

the data). Computing groups working isolated from the clinical questions and acquisition hardware may produce solutions to non-existent problems, or use data that are degraded by poor acquisition technique.

2.7. Early QMRI:

Premature babies were studied with ^{31}P MRS in 1983 (Cady *et al.*, 1983), prompting the measurement of absolute concentration of metabolites in brain (Wray and Tofts, 1986). In 1985 Bakker completed a PhD thesis 'Some exercises in quantitative NMR imaging'. The aim was to 'assess the potential of NMR imaging and spectroscopy with respect to tissue characterisation and evaluation of tissue response to radiotherapy and hypothermia' (Bakker *et al.*, 1984). Quantification was recognised by some radiologists as having a potential role in studying disease (Tofts and du Boulay, 1990):

> Serial measurements in patients and correlation with similar studies in animal models, biopsy results and autopsy material taken together have provided new knowledge about cerebral oedema, water compartmentation, alcoholism and the natural history of multiple sclerosis. There are prospects of using measurement to monitor treatment in other diseases with diffuse brain abnormalities invisible on the usual images.
>
> When making quantitative measurements, the physicist can adopt the paradigm of the scientific instrument designer, who is presented with a sample (the patient) about which he or she wishes to make the most careful, detailed measurements possible, in a non-destructive way, using the infinitely adjustable instrument (the imager). The biological question to be answered and thus the bio-physical feature to be measured need very careful choice.

Quantitative Magnetic Resonance was the subject of a small meeting organised by the UK Institute of Physics and Engineering in Medicine in 1997 at Dundee, Scotland, and it is here that the expression *qMR* was first used. qMRI has now come to denote that part of MR concerned with quantitative measurements, in the same way that fMRI (functional MRI), MRA (MR angiography), MRS (spectroscopy) and qMT (quantitative magnetisation transfer) denote subspecialties of MR.

3. Measurement in medical imaging

Physical quantities can be *intensive* or *extensive*, and when we are considering various properties and manipulations to quantities, it can be helpful to be aware of these differences. An *intensive quantity*[17] can describe a piece of tissue of any size, and it does not alter as the tissue is subdivided (assuming it is uniform). Examples are density, temperature, colour, concentration, magnetisation, membrane permeability, capillary blood volume and perfusion per unit volume of tissue, texture, and the MR parameters proton density, T_1, T_2, the diffusion coefficient of a liquid, and magnetisation transfer. An *extensive quantity* refers to a piece of tissue as a whole, and subdivision reduces (or at least changes) the value of the quantity. Examples are mass, volume, shape, and total blood supply to an organ.

Some intensive quantities, such as metabolite concentration, local blood flow or local permeability, can be expressed either per unit mass of tissue or per unit volume of tissue. Traditionally, physiologists have used the former system, since the mass of a piece of excised tissue is more easily determined that its volume. In qMRI, where the volume of each voxel is well defined, the latter system is more natural. Conversion from per mass to per volume can be achieved by multiplying by the *density of brain*[18] (1.04 g ml^{-1} or 1040 kg m^{-3} for both white and grey matter (Whittall *et al.*, 1997).

3.1. Images, partial volume and maps

Images and maps are terms used to mean different things. An *image* is produced by the MRI scanner, and has an intensity[19] that

---

[17] Intensive (dictionary definition): of or relating to a physical property, measurement etc. that is independent of mass; extensive: a property that is dependent on mass

[18] for example the normal concentration of water in white matter is about 0.690 g water per g tissue (0.690 kg water per kg tissue), equivalent to 0.718 g water per ml tissue (718 kg water per m^3 tissue) (see 1st edition p91).

[19] often called the *signal intensity*, since it is proportional to the signal voltage induced in the RF

depends on a variety of parameters, including some that describe the tissue (e.g. PD, T_1, T_2, and combinations of these), and some that are characteristic of the scanner (e.g. the scanner transmit flip angle and receiver gain). The image consists of a two-dimensional (2D) matrix of numbers stored in a computer (often part of a 3D image dataset). Each location in the matrix is called a *pixel* (picture element), which is typically square and 1-2 mm wide. The image data come from a slice of brain tissue which has been interrogated, or imaged. This slice has a specified thickness (usually 1-5mm), and each pixel in the image in fact derives from a cuboidal box-shaped piece of tissue, called a *voxel*. The first and second dimensions of the voxel are those of the image pixel, and the third dimension is the *slice thickness*. Often the image dataset is three-dimensional, although we can only see a 2D slice through it at any one time.

The interplay between pixels and voxels is subtle. At times when we are thinking of images, pixels are more natural, and in fact the term originates from the science of interpreting images of two-dimensional surfaces (e.g. in robot vision or remote sensing of the earth by satellite). Yet when we are thinking of the cuboids of three-dimensional tissue from which the pixel intensities originate, voxels are more natural, and serve to remind us to think about the tissue, not the image. Slices of voxels are inside the object, whilst surfaces of pixels are outside the object. Some imaging procedures will use very small pixels ('in-plane' resolution), yet set a large slice thickness (in order to retain signal-to-noise ratio). An extreme example would be a voxel of size 0.7 x 0.7 x 5 mm, which appears to have the ability to resolve small structures, yet any structures that do not lie close to the perpendicular to the slice plane would be blurred by the large slice thickness. In this case the voxel would be shaped like a matchstick (i.e. have a large *aspect ratio*); a more appropriate voxel size might be 1.5 x 1.5 x 2.2mm, which has the same volume (and hence signal to noise ratio, for a given imaging time), but is more likely to resolve small structures. Three-dimensional imaging sequences can give us voxels which are *isotropic* (i.e. have the same dimensions all three directions).

coil by precessing magnetisation in that piece of tissue seen in that voxel of the image.

Structures in the brain have very fine detail and very often there are two (or more) types of tissue inside the voxel. The resulting NMR signal from this voxel is simply a combination, or weighted average, of what each individual tissue would give if it filled the whole voxel. Thus if we are trying to measure the T_1 of grey matter, near to CSF (Cerebro-Spinal Fluid) the value measured will be somewhere between that of pure grey matter and pure CSF, depending on the relative proportions of brain tissue and CSF in the voxel. This is called the *partial volume effect*, and is a major source of error when making measurements in brain tissue at locations near to boundaries with other tissue types. The value measured in the tissue is altered by its proximity to another tissue, and the determination of boundaries and of volumes is brought into error. Partial volume errors can be reduced by using smaller voxels, although the price paid is that of a worsening of the signal-to-noise ratio. An inversion pulse before data collection can remove signal from a tissue with a particular T_1 value (as in the FLAIR and STIR sequences, which null the signal from CSF and fat respectively).

A parametric *map* can be calculated from two (or more) images of the same piece of tissue. A simple example would be to collect two images with differing amounts of T_2-weighting. The ratio of these two images then only depends on the tissue parameter T_2, and is independent of scanner parameters (such as transmitter or receiver settings). By calculating this ratio for each pixel, a third matrix, or map, can be formed, which has the appearance of an image (brain structures can be identified), but is conceptually different from an image, in that individual pixel values now have a numerical meaning (such as value of T_2, in milliseconds, at each location in the brain), rather than representing signal intensity on an arbitrary scale.

3.2. Study design

Many studies set out to compare groups of subjects using the classic Double-Blind Randomised Controlled Trial (RCT) design. Typically, a new MR parameter will be evaluated in a particular disease by measuring it in a group of patients and in a group of controls. The controls could be on placebo or another (established) treatment. Other differences between the groups ('confounding

variables') should be removed as much as possible, hence the need for age and gender matching. The scanning should be carried out at the same time, using *interleaved controls*, rather than leaving the controls until the end of patient scanning (when a step change in the measurement procedure could produce an artificial group difference). Some patients may be on treatment which alters the MR results. Matching can be improved by *dynamic matching*, carried out as part of subject recruitment as the study proceeds. Thus if controls are in short supply, but patients plentiful, then each time a control is recruited, a matched patient is selected from the available patients. In placebo-controlled trials, allocation of a patient to the placebo or treated group can be decided at the time of recruitment, to keep the groups matched at all times. Double-blinding[20] is a powerful way of reducing bias in treatment trials. The person giving the treatment, the person making the measurement[21], and the patient are all blinded to whether they are receiving a genuine treatment or a placebo (table 2).

Inexperienced researchers should beware of 'stamp collecting' when 'interesting patients' are studied, almost at random, with no hypothesis or controls[22]. To design high-quality investigations that will be accepted for publication by the best international journals, the investigator should be aware of what work has already been published or presented at international scientific conferences[23].

1	Optimise the precision beforehand
2	talk to a statistician before and after collecting the data
3	collect interleaved control and patient data
4	control for age and gender during subject recruitment
5	inspect the data in scatter plots
6	model the data, including random and systematic error
7	adjust for age and gender during analysis
8	avoid if possible doing t-tests with many comparisons
9	be aware that correlations are hard to interpret
10	give confidence limits on group means and differences

Table 2: good practice in study design and statistics

A *literature search*[24] should be carried out. Studies should not be replicated unless there is a case for confirming the results with a different group of patients. Methodological pitfalls, as illustrated by existing published work, should be identified before the study begins. Some errors (for example the presence of poor reproducibility which would be detected with repeated scanning, or scanning controls after an upgrade, not interleaved with the patients) will irreversibly destroy the value of the data.

Selection of MR parameters requires thought. To acquire all the parameters discussed in this book would require more time than can be fitted into one examination

[20] the double-blind design is not suitable for treatments where the practitioner plays an essential part in the treatment. This is particularly relevant in so-called 'alternative' therapies (e.g. acupuncture, homeopathy, osteopathy, psychotherapy, and reiki). Although a placebo cannot be given, different treatments can be compared. Even in conventional clinical trials, the patient often guesses whether they have a placebo or not from the side-effects, and also those with greater side effects may be more likely to drop out. More research on methodology may be needed to find suitable study designs to overcome these problems.

[21] ideally this includes both the radiographer making the scan, and the observer analysing the MR data.

[22] The term '*hobby researcher*' describes this phenomenon well

[23] See also chapter 2 section 2.2a on two kinds of study: 'fishing expedition' and hypothesis-driven.

[24] for example using PubMed, from the USA National Library of Medicine, available free of charge on-line http://www.ncbi.nlm.nih.gov . From here you can download pdf files of papers (provided you are logged on to an academic website e.g. a university). Usually you can 'search forwards', i.e. see which papers have cited the paper you are looking at (the 'cited by' list). Thus a complete picture of publications on a particular topic can be built up quite quickly and conveniently.

(although as scanners get faster and techniques are optimised, acquisition times have come down). Parameters should be selected according to the biological changes that are expected in the particular disease being studied. Measuring several relevant parameters can be powerful (see chapter 2 section 2.2 e). Mixed-parameters acquisition can address specific questions (for example diffusion weighted spectroscopy, or MT prepared multi-echo measurements. Multiparametric studies are addressed further in chapter 18.

3.3. *Usefulness of an MR parameter*: From a clinical point-of-view, a potential new quantity to characterise brain tissue can be evaluated by considering three factors[25]

Sensitivity: does the quantity alter with disease? Is the False Negative rate low?
Validity: is it relevant to the biological changes that are taking place
Reliability: is it reproducible? Is the False Positive rate low?

Thus the concept of validity (which is absent from a judgement based merely on accuracy and precision) enables the relevance of a metric to be considered. For example, intra-cranial volume could be measured very accurately and precisely, but would be completely irrelevant in most situations. An alternative viewpoint (closely related) is the set of four psychometric properties often used to assess scores: acceptability, reliability, validity and responsiveness (Hobart *et al.*, 2000). The impact of poor reproducibility on the power of a study can be dramatic (figure 1). Appropriate methods for analysing MR data are still under discussion. The clinical metrics are also being scrutinised, and redesigned (Fischer *et al.*, 1999; Hobart *et al.*, 2000). Developments in psychology may be ahead of those used in this field (Krummenauer and Doll, 2000).

4. The future of quantitative MRI

4.1. Technology and methodology

Since the first edition of this book, MRI technology has advanced. The standard field strength has moved from 1.5T to 3T, with 4.7T and 7T machines becoming more common.

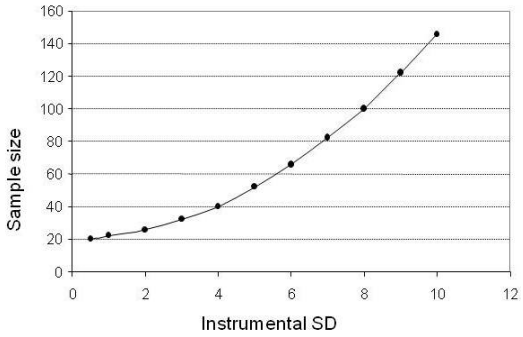


*Figure 1: The effect of instrumental precision (ISD) on the power of a study, and the required sample size. By reducing the ISD, the sample size required is dramatically reduced, with a consequent saving in the cost and duration of the study. This is a simulation based on group comparison between controls (parameter value mean = 100, sd = 3) and the same number of patients (effect size = 5, sd =4.25). Power P=80%, significance level α=0.05, using G*power3*[26].

Major manufacturers offer little below 1.5T for brain imaging. Gradients have improved in both strength and speed, enabling fast 3D acquisition to become standard. RF transmit coils have responded to the higher frequencies by including designs to increase uniformity and reduce SAR. RF receive arrays use multiple coils for improved SNR. The only downsides for qMRI are the need to measure the transmit field B_1^+, and the loss of the reciprocity principle (see chapter 2). Methodologies have continued to advance; this edition has three new topics: advanced diffusion, multinuclear MRS and CEST. Pointers to the future are in chapter 18.

qMRI has five principle aspects:

(i) *concepts* have hardly altered since the 1st edition of this book, just become more clear.

(ii) *MRI physics* above 1.5T is more complex, with the loss of reciprocity.

(iii) *technical advances* continue to alter the environments in which qMRI must be re-implemented.

(iv) *analysis techniques* make a crucial difference to the value of the MRI data[27]

[25] See 1st edition, chapter 12, discussion on by the psychologist N Ramsey in the context of fMRI

[26] G*power3 is established software can be downloaded free-of-charge.

[27] The 1st edition contained 4 chapters on analysis, recognizing that it is at least as important as

(v) the *biological significance*[28] of MR parameters influences what use can be made of qMRI (in particular, can access to the current biology predict the future clinical status?).

Where might these improvements lead? How would we know when no more improvements are worthwhile? We should take time out from the detailed improvements to consider the bigger picture[29]. The concept of the 'Perfect qMRI machine' (see next section) might give a clue. A major improvement in QMRI would come about if there were an international certification scheme for QMRI measurements which have reached the level of the Perfect Machine; a proposal is made below.

4.2. International standardisation and certification.

Standards already exist for measurements in many physical quantities. Readily available machines to measure voltage, body or food weight, and temperature often come with a certificate conforming to the International Standards Organisation[30], guaranteeing a particular performance in terms of total error.

The concept of the 'Perfect Machine' originates in the building of the 200 inch Palomar telescope in the USA in 1933; at the time was the most perfect telescope that could be built[31]. The concept can usefully be applied to an MRI machine used for quantification.

Here it proposed that:

A Perfect Quantitative MRI machine is one that, in making a measurement, contributes no significant extra variation to that which already exists from biological variation.

Various grades of performance can be envisaged, depending on the purpose the measurement. Comparison with normal variation will be the most demanding; comparison with variation within a disease might also be appropriate, depending on the context, and would be less demanding. Here a proposal is made for three levels, each with an appropriate medal[32] (see table 3).

Medal	Target study	Criterion	note
bronze	Group comparison	ISD < 0.3 GSD	(a)
silver	Multicentre study	BCSD < GSD	(b)
gold	Serial study	ISD < 0.3 WSSD	(c)

Table 3 qMRI medals for Perfect Machines: a proposal

Abbreviations:
SD: standard deviation
BSD: biological SD
GSD: group SD
ISD: Instrumental SD,
BCSD: between-centre SD

Notes:
(a) in a group comparison, within-group variation GSD^2 should dominate (i.e. machine variation ISD makes an insignificant contribution to total within-group variation)
(b) the effect of between-centre variation (BCSD) should be less than within group variation
(c) in a serial study, total within-subject variation $WSSD^2$ should dominate (i.e. machine variation ISD makes an insignificant contribution to total within-subject variation).

Bronze medal: In a group comparison, the total variance in each group determines the power and sample size needed (see fig 1). This is the sum of the variance from genuine biological spread (characterised by an SD

acquisition. Spatial registration, shape, texture, volume, atrophy and histograms were considered

[28] The 1st edition had a chapter on the biological significance of MR parameters in multiple sclerosis. Post-mortem studies of tissue can establish a relationship between a qMR parameter (e.g. NAA) and a biological parameter (e.g. neuronal density). Often the relationships are complex and depend on which biological parameters can vary (i.e. the disease context). Spatial registration of the biological specimen and the MR image is crucial.

[29] In Thomas Mann's Death in Venice, the writer is on the Venice beach. He sees the detail, in the foreground: children constructing a sand castle. He turns his gaze to the horizon, empty and infinite. What would it be to be a measurement hero?

[30] E.g. ISO 17025 is the main ISO standard used by testing and calibration laboratories

[31] Building the 200 inch Hale telescope at Palomar, California is described in the book The Perfect Machine: Building the Hale Telescope

[32] Medals are proposed, inspired by the ISMRM use of medals to acknowledge sponsorship at its annual scientific meeting

equal to BSD) and that given by the imperfect machine (characterised by an instrumental SD equal to ISD) i.e. total group variance $GSD^2 = BSD^2 + ISD^2$. Thus if ISD = 0.3 GSD, the contribution of machine variance to the total variance is 9%, and may be considered negligible. This concept allows a Perfect Machine for group comparisons to be specified (table 3). The criteria for each MR parameter would vary; some might be easy to achieve, others might need a long sustained effort. The value of BSD would depend on the kind of subjects considered; in pooling normal values, a correction for age and gender dependence should be applied (treating them as a confounding variable, and standardising all values to a fixed age and gender). The estimates of SD have associated uncertainties, which are significant if the sample size is small (see chapter 3, eqn 2 and figure 6), and these would need to be taken into account when considering if a criterion had been reached.

An example might be the MTR results reported in chapter 3 (figure 8). The stable scanner gave a normal group SD GSD = 0.4 pu, and a measured instrumental SD ISD = 0.15 pu. From these, ISD = 0.375 GSD, and criterion in table 3 (ISD < 0.3 GSD) is not quite satisfied.

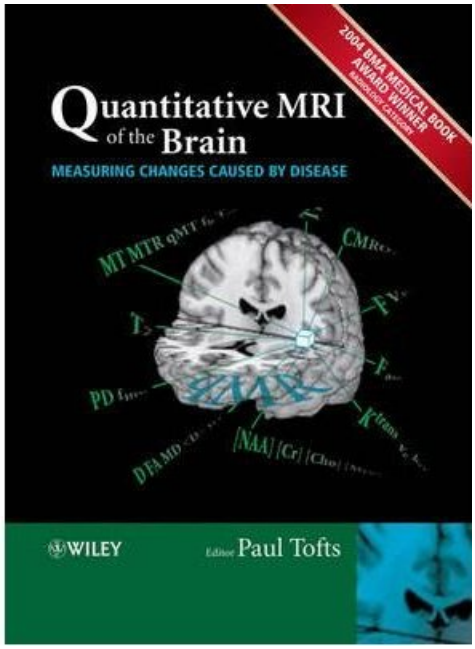
Silver medal: in multicentre studies, inter-centre variation has to be controlled, although some differences can be absorbed by the statistical analysis (provided each subject is always imaged at the same centre). MTR histogram matching using body-coil transmission(Tofts *et al.*, 2006) is probably a perfect silver-medal MTR machine.

Gold medal: in a serial study, instrumental variation can hide subtle within-subject biological changes. The power of a serial study can be limited by such biological variation; often it is small and unknown, and may be extremely hard to measure. Gold medals will be the hardest to obtain; for some MR parameters the gold medal may be impossible. An exception is cerebral blood perfusion, measured by ASL (chapter 16). The natural within-subject variation is large (10-20%) (Parkes *et al.*, 2004) and it might not be difficult to build a Perfect Gold ASL Machine (i.e. one with ISD < 3%). A 2nd example might be in the context of a serial study in relapsing-remitting MS. The within-subject variation in lesion load is highly variable, and perfect gold-medal machines for lesion volume already exist.

Who might administer such a scheme? Award of medals might be determined by the reviewers of a paper submitted to a journal claiming the status, or by an international committee (perhaps sponsored by the ISMRM). Prizes could be awarded (a kind of modern day John Harrison Longitude prize[33]).

The closing words from the 1st edition are still true:

> *Progress towards such automation [of measurement techniques] will take time, and the persistence of John Harrison the clockmaker may enable us to put our work into its historical perspective. We are present at a true technological revolution which is exposing our inner biological workings in ever increasing detail. A few decades ago this was inconceivable; in a few decades' time the techniques will be as routine as measuring the mass of the body.*



[33] Section 2.2

INDEX

REFERENCES:

Bakker CJ, de Graaf CN, van Dijk P. Derivation of quantitative information in NMR imaging: a phantom study. PhysMed Biol 1984; 29(12): 1511-25.

Cady EB, Costello AM, Dawson MJ, Delpy DT, Hope PL, Reynolds EO*, et al.* Non-invasive investigation of cerebral metabolism in newborn infants by phosphorus nuclear magnetic resonance spectroscopy. Lancet 1983; 1(8333): 1059-62.

Filippi M, Dousset V, McFarland HF, Miller DH, Grossman RI. Role of magnetic resonance imaging in the diagnosis and monitoring of multiple sclerosis: consensus report of the White Matter Study Group. J Magn Reson Imaging 2002; 15(5): 499-504.

Filippi M, Grossman RI. MRI techniques to monitor MS evolution: the present and the future. Neurology 2002; 58(8): 1147-53.

Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. Mult Scler 1999; 5(4): 244-50.

Gribbin J. Science: a history 1543-2001. London, UK: Penguin; 2003.

Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. Brain 2000; 123 (Pt 5): 1027-40.

Krummenauer F, Doll G. Statistical methods for the comparison of measurements derived from orthodontic imaging. EurJ Orthod 2000; 22(3): 257-69.

Mallik S, Samson RS, Wheeler-Kingshott CA, Miller DH. Imaging outcomes for trials of remyelination in multiple sclerosis. J Neurol Neurosurg Psychiatry 2014; 85(12): 1396-404.

McFarland HF, Barkhof F, Antel J, Miller DH. The role of MRI as a surrogate outcome measure in multiple sclerosis. MultScler 2002; 8(1): 40-51.

McGowan JC. On the use of quantitative MR imaging. AJNR AmJ Neuroradiol 2001; 22(8): 1451-2.

Miller DH. MRI monitoring of MS in clinical trials. Clin Neurol Neurosurg 2002; 104(3): 236-43.

Parkes LM, Rashid W, Chard DT, Tofts PS. Normal cerebral perfusion measurements using arterial spin labeling: reproducibility, stability, and age and gender effects. Magn Reson Med 2004; 51(4): 736-43.

Sobel D. Longitude: Harper Perennial; 2005.

Sormani MP, Bonzano L, Roccatagliata L, De Stefano N. Magnetic resonance imaging as surrogate for clinical endpoints in multiple sclerosis: data on novel oral drugs. Mult Scler 2011; 17(5): 630-3.

Tofts PS. Quantitative MRI of the brain: measuring changes caused by disease. New York: Wiley; 2003.

Tofts PS, du Boulay EP. Towards quantitative measurements of relaxation times and other parameters in the brain. Neuroradiology 1990; 32(5): 407-15.

Tofts PS, Steens SC, Cercignani M, Admiraal-Behloul F, Hofman PA, van Osch MJ*, et al.* Sources of variation in multi-centre brain MTR histogram studies: body-coil transmission eliminates inter-centre differences. Magma 2006; 19(4): 209-22.

Tofts PS, Wicks DA, Barker GJ. The MRI measurement of NMR and physiological parameters in tissue to study disease process. Prog Clin Biol Res 1991a; 363: 313-25.

Tofts PS, Wicks DAG, Barker GJ. The MRI measurement of NMR and physiological parameters in tissue to study disease process. In: Ortendahl DA, Llacer J, editors. Information Processing in Medical Imaging: Wiley-Liss Inc; 1991b. p. 313-25.

Whittall KP, MacKay AL, Graeb DA, Nugent RA, Li DK, Paty DW. In vivo measurement of T2 distributions and water contents in normal human brain. Magn Reson Med 1997; 37(1): 34-43.

Wray S, Tofts PS. Direct in vivo measurement of absolute metabolite concentrations using 31P nuclear magnetic resonance spectroscopy. Biochim Biophys Acta 1986; 886(3): 399-405.