

Multicentre imaging measurements for oncology and in the brain

^{1,2}P S TOFTS, BA, DPhil and ³D J COLLINS, BA, MInstP

¹Brighton and Sussex Medical School, Brighton, UK, ²Institute of Neurology, University College London, London, UK, and ³Royal Marsden Foundation Trust and Institute of Cancer Research, Belmont, UK

ABSTRACT. Multicentre imaging studies of brain tumours (and other tumour and brain studies) can enable a large group of patients to be studied, yet they present challenging technical problems. Differences between centres can be characterised, understood and minimised by use of phantoms (test objects) and normal control subjects. Normal white matter forms an excellent standard for some MRI parameters (e.g. diffusion or magnetisation transfer) because the normal biological range is low (<2–3%) and the measurements will reflect this, provided the acquisition sequence is controlled. MR phantoms have benefits and they are necessary for some parameters (e.g. tumour volume). Techniques for temperature monitoring and control are given. In a multicentre study or treatment trial, between-centre variation should be minimised. In a cross-sectional study, all groups should be represented at each centre and the effect of centre added as a covariate in the statistical analysis. In a serial study of disease progression or treatment effect, individual patients should receive all of their scans at the same centre; the power is then limited by the within-subject reproducibility. Sources of variation that are generic to any imaging method and analysis parameters include MR sequence mismatch, B₁ errors, CT effective tube potential, region of interest generation and segmentation procedure. Specific tissue parameters are analysed in detail to identify the major sources of variation and the most appropriate phantoms or normal studies. These include dynamic contrast-enhanced and dynamic susceptibility contrast gadolinium imaging, T₁, diffusion, magnetisation transfer, spectroscopy, tumour volume, arterial spin labelling and CT perfusion.

DOI: 10.1259/bjr/74316620

© 2011 The British Institute of Radiology

There have been many approaches to carrying out multicentre imaging studies. A common difficulty is a measurement made at one centre is often not reproducible at another centre and to pool measurements from several centres into a large trial reduces its statistical power. Yet there is a strong imperative to be able to carry out meaningful multicentre studies so clinical drug trials with a large number of subjects can take place in a reasonable time. Analysing the sources of intercentre variation, with insight into the MR physics aspects of the imaging process, and then reducing them where possible has allowed progress to be made. Measuring actual quantities, such as volume, relaxation time or transfer constant, has meant in principle we are able to obtain values independent of the scanner used and the particular way the measurement was carried out.

Multicentre imaging studies are desirable for several reasons. 1. In clinical treatment trials, they are usually the only way of achieving the large number of patients needed to statistically power the study. 2. Measurement of good intercentre agreement demonstrates the imaging technique is good and the results of clinical or scientific research can be applied to other centres. 3. Good intercentre agreement also demonstrates that the physical factors involved in the measurement process are relatively well understood and controlled, and that the

process is relatively reliable and robust and serial measurements through imager upgrades are also likely to be reliable. 4. Pharmaceutical companies will sometimes recruit a large number of clinical centres to engage with many potential buyers of the treatment (although this can actually reduce the power of the study as discussed later).

Multicentre studies have considerable difficulties [1]. These are primarily related to logistics and variance. Logistical problems relate to the interfacing required by the reading centre to the many different imaging centres, which each have a potentially different imaging hardware and protocol, system of controlling and documenting the protocol, and image data format. The resources required increases approximately proportional to the number of imaging centres used. Variance problems relate to the way image measurements of a given subject can vary between centres (the between-centre variance) and how the variance can mask the effects of treatment. A small number of carefully selected and controlled centres, which give high quality data, might give better statistical power than a large number of relatively low quality centres. Minimising between-centre variance (and also within-centre variance) is at the heart of this paper.

In this paper we discuss measurement variance in general terms, including how it affects cross-sectional and serial measurements. The principle sources in MR and CT measurements are identified. Concepts of quality assurance (QA) using normal human controls (healthy volunteers) and phantoms (test objects that seek to simulate

Address correspondence to: Professor Paul Tofts, Brighton and Sussex Medical School, Brighton BN1 9PS, UK. E-mail: p.s.tofts@bms.ac.uk

human tissue) are presented. For individual imaging parameters, the major sources of variance are identified and published measurements are reviewed. Recommendations are given for the design of multicentre studies.

Concepts in multicentre studies

A good ("qualified") biomarker should have three properties [2]: biological relevance to the disease process under study, sensitivity to the disease process and reliability (*i.e.* good reproducibility). Relevance and sensitivity will have been established in single-centre studies. Reproducibility of measurements will have been good at single centres where the initial studies were carried out (to establish the sensitivity). However, at multiple centres this will have to be established again.

In a serial study, to establish treatment effect, each subject will normally be imaged at the same centre for each time point and it is the within-subject variance (*i.e.* reproducibility) measured at a given centre, over the duration of the study, which is important; this is also true for a single-centre study. The between-centre effects constitute a confounding variable; if these effects are small, they can be accounted for by including the centres as a covariate in the statistical analysis [3].

Between-centre variance should also be reasonably controlled, otherwise between-centre effects that alter sensitivity might be suspected and cross terms containing centre and sensitivity (or effect) would be needed in the analysis. In a cross-sectional study, between-centre variance could dominate the power of the study, although the effect of centre as a confounding variable could again be reduced by using centre as a covariate in the statistical analysis, and ensuring each subject group is well represented at each centre. In summary, within-subject variance should already be controlled in any single-centre study; if the study is to be multicentre then between-centre variance should also be controlled.

The within-centre variance for a subject is important; this is most straightforwardly measured using the Bland-Altman analysis method [4, 5]. Repeated measurements are made (usually in pairs) over a set of subjects (typically 5–20) to establish the difference between repeats and whether this depends on the mean value of the parameter being estimated. This variance can then be used directly in a power calculation. This variance usually depends (sometimes quite dramatically) on the time period over which it is measured. A repeated image data collection, while the subject remains on the scanning couch, will only be vulnerable to image noise and perhaps movement. Conversely, a pair of measurements separated by several years could be vulnerable to factors such as imager hardware upgrade, change of radiographer and change of image analysis procedure.

Protocol-matching is the most straightforward way of reducing between-centre differences. For example, in reporting multiple sclerosis (MS) hypointense lesions from T_2 weighted MRI, differences in qualitative image appearance could be reduced by standardising the acquisition parameters thought to be relevant (repetition time (TR), echo time (TE), slice thickness, type of radio-frequency (RF) coil etc.) and the analysis procedure (by defining rules) [6–9]. The MAGNIMS (magnetic

resonance in multiple sclerosis) study group defined many guidelines for multicentre MR studies in MS [10, 11]. At times this involved MR physicists travelling to different sites within Europe, sometimes with phantoms, to improve the protocol matching [8]. Differences between imaging hardware produced by different manufacturers (vendors) may prevent identical protocols being used at every site. Alternatively, acquisition parameters may be compromised, giving a protocol common to all sites, but this will be less than optimal for most sites.

Quantitative imaging maps (derived from source image data and a common post-processing software package) may further reduce between-centre differences, since parameter maps of particular quantities are calculated, which in principle are independent of the particular imaging protocol used. For example, apparent diffusion coefficient (ADC) maps are approximately independent of the particular imaging parameters used (*e.g.* TE, gradient strength, b -value and static field B_0). In some cases there may be a conflict between the requirements to minimise within- and between-centre variances. For example, spectroscopic metabolite peak ratios usually have better within-subject reproducibility than absolute metabolite concentrations do, since the latter quantity involves input from more potentially variable sources related to the instrument calibration. Conversely, the between-centre performance of absolute concentrations is often better, since more centre-specific factors have been taken in to account in the calculation.

Sources of variation

Biological variation and the perfect instrument

The magnitude of genuine biological variation in a subject is usually unknown. This is the variance measured with an instrument that has perfect reproducibility (or at least negligible variance of its own) and represents how much the biology of a tissue varies over a short period of time (say 1 day). For example, it is likely that blood flow or blood volume do change over short periods of time (*e.g.* in response to cardiac output) and that tumour volume probably does not change. Other factors that could alter tissue biology in the short term include the state of hydration and menstrual cycle. These all appear as random effects, which can mask the effect of the treatment.

If the instrumental variance is well below this biological variance, then the MR imager constitutes a "perfect instrument"; the power of a study is no longer limited by instrumental imperfections. Thus, we can aspire to this situation. In MS, the visible lesions vary enormously with time (by maybe 50%). This is the phenomenon of relapse and remission and therefore a perfect instrument can easily be achieved (the requirements on instrumental reproducibility are very lax). Conversely, the changes observed in brain tissue of normal appearance in MS are very small (typically 3%), instrumental performance probably limits what can currently be observed and there is a premium on improving instrumental performance. The magnitude of instrumental variance compared with the biological changes seen as a result of disease determines the importance of instrumental imperfections for each tissue parameter and the amount of effort that

should be expended in reducing such imperfections. This effort might include setting up and maintaining quality assurance (QA) at each centre and improving the acquisition procedure. For a technically demanding parameter this process might be done at only a few high-quality centres, while for a parameter for which the instruments are already perfect, many more centres could easily be included in the study.

The process of imaging, whether for qualitative reporting or for quantitative measurement, can be conceptually split into the sub-processes of data collection and image analysis. This is convenient for identifying the sources of variation, since the analysis process for a given dataset can usually be repeated (*e.g.* for a Bland-Altman analysis), while data collection often can not be repeated (particularly if it involves administration of contrast agent or the use of ionising radiation).

Image data collection

All image data collection procedures are potentially vulnerable to the following sources of variation [3, 12]:

- differences in acquisition parameters
- patient repositioning errors between examinations
- patient gross movement during the scan (*e.g.* moving the head slightly); this can be minimised by careful attention to patient comfort
- involuntary patient movement during the scan:
 - respiratory motion
 - cardiac motion and ensuing blood vessel and ventricular pulsation
 - gut motion
- image noise
- upgrades to imaging hardware or acquisition software.

MR acquisition is also vulnerable to the following effects:

- The pre-scan procedure to set imaging parameters (*e.g.* flip angle (FA) and shimming) may be variable.
- Changes in static field strength: some tissue parameters are, in principal, independent of field strength (K^{trans} , diffusivity, volume, metabolite concentration, blood volume and perfusion). (However, an altered measurement protocol and the increased signal-to-noise ratio at 3 T compared with 1.5 T may be evident). Other parameters are certainly dependent on field strength (*e.g.* T_1 , magnetisation transfer).
- RF coil architecture, which affects the B_1 and hence FA distribution (whether head or body coil excitation is used). Body coil excitation is preferable [12], although there can still be differences between vendors, particularly at 3 T.

CT acquisition is also vulnerable to changes in X-ray tube output. Although a daily calibration for water CT number is usually carried out, changes in the beam energy spectrum will not be compensated for. Any resulting change in beam effective energy will alter the measured CT number of tissues containing significant amounts of high atomic number elements.

Image analysis

Image analysis results often depend on the observer carrying out the procedure. Sources of variation can include:

- image viewing conditions (display window, image magnification, viewing distance, ambient lighting)
- the set of rules used to define (segment) a region of interest (ROI) around an area of abnormality (*e.g.* tumour outline) and to generate a histogram (if used) [12]
- mode of operation of analysis software
- changes in analysis software over time.

In multicentre trials, the trend is towards using a single centre for analysis; (the clinical research organisation). If multiple tissue parameters are being evaluated, then distinct centres are often used for each parameter to spread the analysis load. Observers are trained to a set of well-defined rules and procedures and their performance in repeated analysis is monitored. With manual analysis of serial data, there may be long-term drift [1]. Best practice is to repeat the analysis each time there is a new time point for a given subject in addition to continuously monitoring the QA results. Software for semi-automated analysis will reduce variance compared with a manual procedure. An entirely automatic procedure must be completely reproducible for the same image dataset. In multicentre scientific studies, several centres may wish to carry out analysis and then these sources of variation would be identified and controlled as much as possible, particularly by moving towards using a single software analysis system (*e.g.* in IDL or java languages), which can be run on many platforms; this also has the benefit that ROIs are all generated in a common format, with a common naming convention.

Batch processing files (*e.g.* in XML language) are often used to drive the analysis process; this file records the entire processing procedure in a way that can be read by a human [13]. The file might include information on the analysis model used, the kind of T_1 calculation and the software version used. This produces an audit trail, as required for FDA approved clinical trials and any result can be exactly reproduced at a later date.

Quality assurance

Accuracy and precision

Accuracy and precision are determined in QA procedures that determine the quality of the imaging process. They can be cross-sectional (across several centres) to estimate between-centre effects, or serial (at a single centre) to monitor stability over time. They can be used to monitor two quantities. Accuracy (*i.e.* closeness to the truth or systematic error) can be found using a phantom with known properties (*e.g.* volume or diffusion coefficient). Precision (stability, reproducibility, test-retest difference or random error) can be found with phantoms (which often need temperature

control) or human subjects (provided they are known to be stable over the period of the repeated measurement) (Table 1). Stability can be measured over various time periods, from a few minutes (with a repeated scan keeping the subject on the scanner bed) to several years (through scanner upgrades). Tumours usually change quickly enough so that long-term measurement stability is not important. If accuracy can be established at each centre (e.g. if measured values of a tumour volume standard, or of normal white matter T_1 , are close to the known values) and if the precision (i.e. within-centre reproducibility) is good, then the between-centre differences will be small and need not be measured explicitly.

Control subjects (healthy volunteers)

Human control subjects have three benefits over phantoms (at least for MR studies). They can be an almost completely realistic simulation of the clinical measurement process in a multicentre study (depending on the tissue parameter). The white matter of healthy controls (frontal or parietal lobes) often has a narrow range of tissue parameters and can be used at different centres without the need to transport the subjects between centres. The normal range can be estimated from published high quality measurements made on a number of controls at a single centre (ideally where within-subject reproducibility has also been measured) and age and gender may need to be controlled for. Some recent single-centre studies of normal subjects report encouragingly narrow ranges of values, with measured normal coefficients of variation (CV) for magnetisation transfer ratio (MTR) and diffusivity down to about 1–3% (see Tables 2 and 3). Reported values usually include the effect of instrument variation (although this could in principle be removed with an analysis of variance); therefore, the lower published values probably approach an upper limit to the true biological variation (which could be considerably lower).

Secondly, homeostasis provides inbuilt temperature control and the demands of temperature stabilisation are bypassed. Thirdly, human controls are often more readily available than phantoms.

Disadvantages to using human subjects include: i) a lack of stability over time for tumour-related parameters in patients (e.g. for K^{trans} and volume); thus the stability of a measurement procedure cannot be evaluated. ii) Imaging humans may be more demanding of resources than imaging phantoms, depending on the operating policy of the imaging centre. iii) If the imaging procedure includes incidental hazards, such as contrast agent injection or the use of ionising radiation, then the availability of normal controls and the number of repeated exams will be limited by ethical constraints.

Phantoms

Phantoms exist for some of the tissue properties that can be measured in tumours, but not all [2]. They have

the benefit (over human control measurements) of i) being convenient to scan as many times as required without any ethical constraints (particularly important for CT), ii) having known physical properties and iii) being relatively easy to transport between centres.

Their potential disadvantages include: i) there may be a lack of realism compared with in-vivo measurements, some quantities (e.g. blood flow) are extremely difficult to simulate in a phantom and the measurement process, even for a simple quantity like volume, does not replicate the complexity of the *in vivo* process. ii) The MR properties may vary with time, caused by either a temperature dependence of the parameter under study (e.g. diffusion) or by instability of the material over time (e.g. caused by fungal attack, chemical decay, or gel breakdown). For liquids or gels, failure of the container cap seal can lead to evaporation, oxidation or ingress of water vapour. iii) The time and expertise required to manufacture phantoms may be prohibitive at some centres. These phantoms are designed to directly monitor imager performance in measuring tissue parameters; they are distinct from those designed to measure instrumental parameters such as slice thickness or image noise [14].

Temperature control

Temperature control of MR phantoms is now recognised as being crucial if imager stability is to be measured precisely (i.e. within <1%). The temperature coefficient of MR diffusion and T_1 phantoms is about 2.5% per °C [2]. As a result, the temperature must be

Table 1. Recommended quality assurance procedures for each tissue parameter using phantoms, normal subjects and tumours. Phantom measurements can be used in an initial evaluation where they exist [T_1 , diffusion, magnetisation transfer ratio (MTR), spectroscopy and volume]. Measurements in normal brain give precision (from repeated measurements) and accuracy (by comparison with published values) in T_1 , diffusion, MTR, spectroscopy and arterial spin labelling (ASL), with more realism than phantoms. Measurements in tumours are particularly important where no phantom or normal tissue measurements are appropriate (K^{trans} and volume)

Quantity	Phantom	Normal brain	Tumour
T_1 weighted DCE ^a	None	No ^c	Yes
T_1	Yes	Yes	
T_2 weighted DSC ^b	None	No ^c	Yes
Diffusion	Yes ^d	Yes	
MTR	Yes	Yes ^e	
Spectroscopy	Yes	Yes	
Volume	Yes	No	Yes
ASL	None	Yes ^f	
CT perfusion	None	No ^g	Yes

^aFor K^{trans} .

^bFor blood flow and volume (rCBF and rCBV).

^cIt may sometimes be possible to give gadolinium to normal controls.

^dApparent diffusion coefficient or mean diffusivity only (not tensor).

^eValues depend on sequence.

^fNB large normal variation [96].

^gIt may sometimes possible to measure CT perfusion on patients with normal brain.

monitored and controlled to within about 0.2°C to confirm a stability or precision of 0.5% in these parameters. Storage of phantoms in the magnet room, at a temperature as close to that of the magnet bore as possible, reduces phantom temperature changes during scanning. RF power deposition in phantoms can usually be ignored unless they have been deliberately made electrically conducting by the addition of NaCl. This can be confirmed by monitoring the phantom temperature before and after scanning.

Measurements near to body temperature can be made by placing the phantom and containers of warm water (or saline) inside an insulating container. The temperature of the phantom can be monitored by using thermocouples (or the more expensive and brittle optical probes) attached to the surface of the container provided there is plenty of thermal insulation around the container. Alternatively, the thermocouples can be placed inside the phantom (if liquid), although this involves breaking into the container. A thin T-type thermocouple has signal dropout limited to within about 3 mm of the tip [15].

Thermal insulation

Thermal insulation can conveniently be achieved by re-using packaging made of expanded polystyrene, or can be custom-made for better performance. Thermal insulation slabs (approximately 100 mm thick) used in the construction industry for cavity wall insulation are readily available and perform well, although they may have a layer of aluminium foil attached, which has to be painstakingly removed first. Expanded polystyrene is commonly available and cheap. Expanded urethane or phenolic foam, although slightly more expensive, has considerably lower values of thermal conductivity [16] (approximately 60% of the value for expanded polystyrene) and the fine mechanical structure makes it easier to machine. Approximate values for thermal conductivity (k) (data from an internet search) are as follows: mineral wool, $0.041 \text{ W m}^{-1} \text{ K}^{-1}$; expanded polystyrene, $0.037 \text{ W m}^{-1} \text{ K}^{-1}$; rigid urethane foam, $0.023 \text{ W m}^{-1} \text{ K}^{-1}$ and rigid phenolic foam, $0.022 \text{ W m}^{-1} \text{ K}^{-1}$. For comparison, air has $k=0.024 \text{ W m}^{-1} \text{ K}^{-1}$. One of the authors (PT) has successfully used Kingspan Kooltherm K8 phenolic 50 mm thick cavity board (Kingspan Insulation Ltd, Herefordshire, UK), which is easily available in the UK from builders merchants. Thermal time constants greater than 10 h can be obtained by using a 5 cm bottle in a head coil, and filling all the available space with insulation. Machining is straightforward using a bandsaw and slow drill bit, although care should be taken to avoid any small ferromagnetic chards becoming embedded in the phantom.

Provided the phantom temperature has been measured during QA measurements, then the phantom parameter values (e.g. ADC) can be converted to a standard temperature for realistic comparison with those made at other times (which will inevitably have been made at slightly different temperatures) [17]. In some cases the phantom can be made using available liquids from a chemical supplier [17, 18]; this has the convenience of being able to leave the liquids sealed in the delivery bottle (although slightly magnetic labels might have to be removed).

Reanalysis

Reanalysis of acquired image data (both human and phantom) estimates how much the analysis process contributes to variation. Generally this process can be refined and semi-automated so the majority of the measurement variance is derived from acquisition. Analysis of the same image data at different centres can identify between-centre effects derived from analysis rather than acquisition.

Specific QA procedures for each tissue parameter

Qualitative weighted images

Qualitative weighted images (e.g. qualitative T_2 weighted images) are often used in multicentre studies because they are straightforward to implement at each centre. Lesion dimensions can be estimated manually. By controlling the acquisition parameters (e.g. TR, TE, TI, voxel dimensions, etc), the display parameters (e.g. greyscale windowing, image magnification, ambient lighting etc.) and the rules for radiological analysis (e.g. how to identify the border of a tumour and maximising the use of automated analysis software), then a reasonable between-centre agreement can often be achieved [6] without the need for a large amount of technical input (Table 3).

T_1 weighted dynamic contrast-enhanced imaging

T_1 weighted dynamic contrast-enhanced (DCE) measurements may be used to calculate semi-quantitative parameters such as the initial area under the gadolinium curve (IAUGC) [19] or K^{trans} [20]. Within-centre variance may be better for IAUGC since variation from sources such as the arterial input function (AIF) may be reduced. Conversely, K^{trans} may be successful at removing the effects of varying AIF at different centres and result in better multicentre performance. The major sources of variance are likely to be: the injection procedure (a power injector makes this more reproducible); the image acquisition procedure (this determines spatial and temporal resolution and image noise); the way that T_{10} [the tumour T_1 before injection of gadolinium(Gd)] is measured, if at all; the way that the AIF is derived from a ROI over a major blood vessel (the descending aorta or a carotid can be imaged using a different coil from that used for the brain); the way that ROIs are defined on the tumour; and the modelling procedure used to estimate IAUGC or K^{trans} .

There are currently no realistic QA phantoms for T_1 weighted DCE, although renal dialysis cartridges have been used by some. Patients with tumours are unlikely to be stable, although short-term studies can be made within- or even between-centres. There have been few studies of T_1 weighted DCE variation, largely owing to ethical problems around administering repeated doses of contrast agent [21–23].

In the future, the use of a pre-bolus technique [24, 25] where a small dose (typically one-tenth of the standard dose) is given to determine the AIF is likely to be adopted. The accuracy of the AIF estimation is improved since the dynamic range of the signal is less and inflow

effects can be avoided by using an appropriate sequence; also its reproducibility can be determined.

T_1 mapping

T_1 weighted DCE when used quantitatively is inherently a dynamic T_1 determination and the factors that control T_1 mapping accuracy are also influential in T_1 weighted DCE. The tumour T_1 must be measured before the injection of Gd (*i.e.* T_{10}). A convenient way to establish the T_1 weighted sequence is behaving as expected is to also collect a proton density weighted image (with reduced flip angle), estimate normal white matter T_1 from the ratio of these images [26], and compare with published values. The T_1 sensitivity of the spoiled gradient echo sequence is dependent on flip angle. Therefore any B_1 errors can make the T_{10} determination inaccurate unless a B_1 map is made [27] and they will also affect the amplitude of the post-Gd signal.

Published values of T_1 in normal human white matter show some between-centre variation (probably because of B_1 effects and ROI placement) and there are few large studies of normal variation. The measured variation (Table 2) is quite small (CV=4–6%), and the intrinsic biological variation may be much smaller since diffusion and MTR have even smaller measured CVs (Tables 2 and 4). At 3 T, white matter values [30, 31] are about 1050–1080 ms. Given the vulnerability to B_1 errors, which can easily be different in phantoms and human brain, human brain QA is advised for T_1 .

Phantoms for T_1 are well established [2]. These can be aqueous solutions of paramagnetic ions made in-house or purchased from commercial sources, such as the Eurospin set (Diagnostic Sonar, Livingstone, UK).

Multicentre studies on T_1 are rare. Deoni et al [32] measured a mean value in normal white matter at 1.5 T of 662 ms at 3 centres: standard deviation (SD)=44 ms, CV=7%. The between-centre variation was comparable with the within-centre variation.

Interest in T_1 measurement methods may increase because of its contribution to DCE and as a surrogate for water content. If it does, then more standardised methodology will evolve, including B_1 correction, particularly for 3 T.

Table 2. Normal range of T_1 values in white matter

Study	Year	CV ^a (%)	n ^b	Mean (ms)	SD (ms)
Stevenson et al [28]	2000	5	40	666	36 ^c
Rutgers et al [29]	2002	6	15	681	40
Ethofer et al [30]	2003	4	8	770 ^d	30

CV, coefficient of variation; SD, standard deviation.

^aCoefficient of variation=SD/mean.

^bSample size.

^cEstimated from box-plot in figure.

^dUsed spectroscopic technique; probably some cerebrospinal fluid or grey matter contamination.

T_2 weighted dynamic susceptibility contrast for blood flow (perfusion) and volume

The sources of variation are very similar to those encountered in T_1 weighted DCE (see T_1 weighted dynamic contrast-enhanced imaging section). Relative measurements (*i.e.* ratio to contralateral normal-appearing tissue) are often more reliable than absolute measurements. There is very little literature on dynamic susceptibility contrast (DSC) QA or multicentre studies [33].

Diffusion

Diffusion measurements are potentially quite reliable because they are insensitive to B_1 errors. Gradient values are usually accurate to within 0.5% in order to give correct object dimensions and ADC values should, in principle, be accurate to within 1% [18]. The implementation of the gradient scheme, in which a range of b -values are produced in a range of directions, is often provided by the manufacturer. Artefacts can cause problems, depending on the sequence. These may occur as a result of subject movement or susceptibility induced signal dropout in the EPI sequence or may originate in the scanner itself (*e.g.* if the gradients cannot be switched accurately enough). In a situation where intrinsic ADC values are known to be accurate (*e.g.* in large regions of liquid or normal brain), there may still be large variations when lesion or tumour ADC values are measured due to variations in the ROI generation procedure [34].

Mean diffusivity can be conveniently validated using normal white matter, which has measured values in the range 0.69–0.93 $\times 10^{-9}$ $\text{m}^2 \text{s}^{-1}$ [35–42]. The measured value depends on the b -value, the diffusion time Δ [43, 44] and the sizes of the ROI and voxel. Large ROIs or voxels may produce more cerebrospinal fluid (CSF) or grey matter contamination. There is a small increase with age [45]. In white matter, low normal ranges (3–5%) for measured values have been reported and the intrinsic normal biological variation is almost certainly below this (Table 3). For within subject variation measured by repeated scans, CVs of 2.5% [46] and 2.6% [47] have been reported.

Table 3. Normal range of mean diffusivity values in white matter

Study	Year	CV ^a (%)	n	Mean (10^{-9} $\text{m}^2 \text{s}^{-1}$)	SD (10^{-9} $\text{m}^2 \text{s}^{-1}$)
Cercignani et al [39]	2001	5	20	0.93 ^b	0.04
Emmer et al [42]	2006	4	12	0.84	0.03
Zhang et al [40]	2007	5 ^c	29	0.69	0.04
Welsh et al [41]	2007	3	21	0.73	0.02

CV, coefficient of variation; SD, standard deviation.

^aCoefficient of variation=SD/mean.

^bSome cerebrospinal fluid contamination.

^cCV=2% for whole brain.

Phantom measurements can be made with alkanes [17, 18, 48] or other organic liquids [48], which are available with diffusion coefficient (*i.e.* ADC) values in the range of brain tissue. Alternative liquids include sucrose solution [49, 50] or iced water [107]. At room temperature water has an ADC value [48, 51, 52] three times that of normal white matter so the signal is considerably lower. Gels may be non-uniform and unstable.

Validation of fractional anisotropy (FA) or full tensor measurements is much harder. In humans, the corpus callosum has anisotropic diffusion of FA=0.7 [53, 54] and is suitable for an approximate comparison. However, the measured FA value will depend on a variety of factors, including the voxel dimension [55] (larger voxels can average out some of the anisotropy if fibres are not exactly parallel), the amount of image noise [56] and the precise size and location of the ROI (since partial volume effects are quite strong); large within-centre variations have been reported [57]. Some anisotropic phantoms are under development [58, 59], although these are currently quite complex to manufacture and are quite small (up to about 2 cm in diameter).

Two multicentre studies of normal brain tissue [60, 61] found differences of 4–9% and 2–6%, and considerable within-site variation.

In the future, the use of liquid phantoms or normal white matter will become routine and anisotropic phantoms will probably become commercially available. Most scanner faults can probably be identified using liquids, so the added value of using an isotropic phantom may be small.

Magnetisation transfer ratio

The largest source of variation in measuring MTR values are sequence and pulse differences [62], and B₁ factors [12]. MTR values in normal white matter are in the range 30–60 pu (percent units), depending crucially on the sequence used [8, 12, 62, 63]. A common sequence must be chosen before any multicentre comparisons can be made. The normal range is low, below 2%, perhaps partly related to white and grey matter having similar MTR values (Table 4).

A phantom with a realistically high value of MTR can be made from bovine serum albumin [65] (agarose gel does not have enough bound protons) [2]. Azide should be added as a preservative.

In a multicentre study, by implementing the same pulse sequence at several different centres with different manufacturers, a common sequence (EuroMT sequence)

gave almost identical MTR values in normal white matter using transmit/receive head coils [8]. A more recent two-centre study of MTR histograms in normal brain, using two different manufacturers, showed that by controlling the pulse sequence, B₁ factors (using body coil excitation) and the image segmentation procedure, the between-centre variation could be completely eliminated [12].

If manufacturers move to providing more choice in setting the values of the pulse parameters in MTR sequences, then centres can harmonise their data collection procedures. The MT pulse shape, amplitude, offset frequency and TR all have to be controlled. With the flexibility to acquire images at several values of amplitude and offset frequency, a more quantitative MT acquisition procedure is possible, where the concentration of bound protons can be measured [66–68]. More details are given in the MT section in the multi-author techniques article [105].

Spectroscopy

¹H magnetic resonance spectroscopy (¹H MRS) has been proven to be useful in the grading and assessment of therapies in brain tumours [69]. It is also increasingly used in the guidance of radiotherapy treatments [70]. Sources of variation include the pulse sequence and the procedures for shimming, for setting B₁ over the volume of collection and for water suppression (and the resulting efficiency of each process).

Normal values for absolute metabolite concentrations (*e.g.* N-acetylaspartate) are available [71], although institutional units are often used which contain systematic centre effects (such as T₂ dependence); metabolite ratios are often more reliable [72].

To establish a multicentre trial it is essential to develop a number of QA procedures to include phantoms, measurement protocols and post-processing software. Several publications have detailed MRS phantom designs, which all have a common design template [73]. This consists of a larger outer volume containing a smaller internal volume. The containers are normally manufactured using Perspex (acrylic or Plexiglas). The two volumes hold different chemical solutions to enable the discrimination of signal sources under different measurement conditions. MRS localised measurements are normally performed on these phantoms to establish the following: localisation accuracy, efficiency of localisation, degree of contamination, linearity and magnitude of artefacts (MRS line shape distortions). Additional

Table 4. Normal range of MTR values in white matter

Study	Year	CV ^a (%)	n	Mean (pu) ^b	SD (pu)
Silver et al [63]	1997	1.9	41	39.5	0.76 ^c
Davies et al [64]	2005	1.0	19	38.4	0.4
Tofts et al [12]	2006	1.6	10	37.3 ^d	0.6

CV, coefficient of variation; SD, standard deviation.

^aCoefficient of variation=SD/mean.

^bMTR values are not comparable (different sequences).

^cSEM=0.17 pu; four samples each n=20 or 21; estimated SD=0.76 pu.

^dPeak location values in white matter histograms.

measurements include assessments of water line-widths following B_0 optimisation (shimming) in various phantom positions and water suppression efficiency [74, 75]. The MRS measurement sequences used for these tests are available on all of the major vendor platforms. Typically MRS measurements are performed with the standard head coil using single voxel PRESS measurements with a TE in the order of 30 ms, alternatively, single voxel STEAM measurements with a TE of 20 ms are used. Similar measurements can be performed using multivoxel chemical shift imaging measurements, although there are more variations in the vendor implementation of these techniques, which may make direct comparisons more difficult.

Having established a measurement QA protocol, it is important to perform the QA measurements routinely and following any hardware or software upgrades. Following phantom measurements, it is important to acquire sample volunteer MRS spectra to enable expert spectroscopists to evaluate the data before beginning the trial. It is common practice in multicentre studies to establish a single centre for the evaluation of data. This is particularly true with MRS data as it is still a relatively new measurement technique. It is important to evaluate MRS data shortly after acquisition by appropriate personnel using post-processing software that is common across all the participating sites.

Considerable multinational efforts have been made to establish multicentre trials of MRS for the evaluation and classification of brain tumours [76]. Results in normal tissue have been relatively disappointing for spectroscopy [77, 78], although within centre variation can be good (approximately 4%) [72, 79, 80].

Currently there are differences in the vendors' post processing tools for MRS data. To establish common software platform for analysis at each site would require the installation of an accepted standard software package such as MRUI (<http://www.mrui.uab.es>) [81], Tarquin [106] (<http://tarquin.sourceforge.net/index.php>) or LCModel [82] (<http://s-provencher.com/pages/lcmodel.shtml>). Currently there is an ongoing international development to establish automatic pattern recognition tools for classification brain MRS data [76]. Future multicentre trials may be able to both utilise and support this activity.

Tumour volume

There is an increasing interest in measuring tumour volume [83, 84] rather than just making a one-dimensional measurement, such as RECIST [85, 86]. Our (unpublished) results on FLAIR [87] images (5 mm 2D slices) of gliomas show that good accuracy can usually be obtained and the reproducibility is limited by how well the outline of the tumour can be delineated.

Moving patients between different scanners is generally not practical; however, the precision of the tumour segmentation process can be studied by repeated analysis at the same or different sites. In an unpublished study by one of the authors (PT), clinical images could be analysed with a precision of $SD=3.8$ ml ($CV=3.4\%$). Some tumours had indistinct boundaries, which limited precision. A simple phantom was made using bottles of water doped to give realistic values of T_1 [2]. A variety of volumes were

used to cover the range of tumour volumes expected (typically 40–200 ml) and the volume of water was determined by weighing. Scanning in oblique planes to give realistic partial volume effects [88] and adding noise to the images produced a set of realistic images from which accuracy was estimated. Absolute volume in millilitres could be measured to within 3%, the analysis precision was $SD=0.64$ ml ($CV=0.48\%$) and the overall measurement precision was 1.4 ml (1.3%).

Much work has been published on measuring the total lesion volume in MS [6, 7, 9]. Standardising the imaging parameters within a range of values and agreeing rules on the lesion segmentation process reduced within-centre variation to a level that is insignificant compared with the natural variation in the disease process.

In the future there will be more clarification regarding when volume measurements provide added clinical value over one-dimensional measurements in research and clinical environments. As manufacturers provide better techniques for volume measurement this balance may shift.

Cerebral blood flow measured using arterial spin labelling

Arterial spin labelling (ASL) is a quantitative technique for cerebral blood flow (CBF) measurement; however, there are large sources of variation. Numerous acquisition methods and analysis models exist, which have produced a large range of published CBF values. To produce quantitative CBF maps the acquisition sequence and needs to take into account corrections for large vessel flow and arrival time delays and must provide a measurement of the equilibrium magnetisation of arterial blood [89]. Various approaches are provided for these three issues. Firstly, arrival time delays may be corrected by using a multi-inversion time sequence and fitting for arrival time in the analysis model [90]. Alternatively, simply waiting for a longer delay time to allow labelled spins to fully enter the imaging region can reduce the sensitivity to arrival time [91]. The Q2TIPS technique further cuts off the tail end of the perfusing bolus to enhance this effect [92]. Large vessel contamination can be avoided by the addition of dipolar diffusion gradients to dephase the high flow signal in large vessels [93]. Others simply increase the delay time to allow signal in large vessels to exit. Finally, the equilibrium magnetisation of arterial blood can be measured from the equilibrium tissue magnetisation, either whole brain [94], white matter [92] or from CSF [90]. This can be corrected by an appropriate blood:tissue partition coefficient.

A range of analysis models also exist, *e.g.* single tissue and blood compartment models and two-compartment models [89]. The choice of a particular model will depend on the data acquisition technique. If a short delay time between labelling and acquisition is chosen, a single blood-compartment may be appropriate; whereas a longer delay time may need a two-compartment model as some of the labelled spins begin to cross the capillary wall [95].

There is little consensus concerning which of the data acquisition approaches and analysis models are superior.

Each approach will undoubtedly give slightly different CBF values and therefore, the adoption of the same acquisition and analysis approach for multicentre studies is important. This is complicated by the fact that scanners produced by different manufacturers may not have the equivalent ASL sequences available. It may be that the best approach in terms of standardisation is to adopt a simple sequence, such as FAIR, to match acquisition parameters and then use the same analysis model.

At present there are no realistic perfusion phantoms, while normal values show considerable natural within-subject variation, according to the physiological state of the subject [96]. These two factors limit how well the between-centre performance can be measured.

The first large multicentre study used centres with identical scanners. The study involved 199 participants in 22 centres using a Phillips 3T system [97]. This showed reasonable reproducibility with a coefficient of variation of 13% between repeated scans.

In the future the provision of hardware, imaging sequences and analysis software by the manufacturers may converge and a realistic phantom may become available.

CT perfusion

Multidetector CT (MDCT) systems are capable of performing contrast-enhanced dynamic CT (d-CT) [98, 99]. Establishing a multicentre study using d-CT will require the usual definition of a standardised imaging protocol and contrast delivery that can be implemented across all vendor platforms with the added requirement of ensuring the total radiation dose will not become unacceptable if the subject is scanned several times. Particular attention needs to be given to the reduction of skin dose and eye lens during d-CT [100, 101]. This will require the support of an experienced medical physicist capable of making this determination on all of the scanners used in the trial. d-CT protocols normally image a few slices with high temporal resolution (1 s) during the initial first pass of the contrast bolus (50–60 s) and then acquire data at a lower temporal resolution (10–15 s) for a further 2–3 min.

QA procedures will require the use of a multicompartamental iodine phantom to establish the linearity of the Hounsfield number with iodine concentration for the resulting d-CT measurement parameters in addition to the routine vendor specific QA procedures. QA measurements should be made following all software and hardware changes in addition to the daily and weekly routine procedures.

All equipment vendors now provide post-processing software using different pharmacokinetic models for the analysis of d-CT data. However, in several papers a lack of concordance in the results obtained from post-processing the same patient data with different software indicate only one type of post-processing software should be used, ideally in the single centre responsible for the evaluation of data for the trial [102]. Site specific patient data processing should be performed immediately after data acquisition using available software to ensure the studies are of sufficient quality for further evaluation.

Despite the widespread availability MDCT systems there appears to be no documented multicentre study in a neuro-oncological setting. However, there are numerous examples of multicentre CT studies in other radiological settings.

Discussion

One of the first quantities that clinicians attempted to measure in an MRI multicentre trial was the lesion load (*i.e.* total visible lesion volume) in MS. This was carried out by the MAGNIMS group, which was originally funded by the European Union [6, 7, 10]. The intercentre effects were initially large. The sources of discrepancy were identified by bringing together experienced neurologists from several centres to report the same images at the same time and in the same room. They agreed on a set of rules for equivocal cases. The within-subject variation in lesion load is relatively large. It can vary by 5–10% over time as the disease produces relapse and remission. The variations arising from between-centre and other effects are relatively unimportant provided they are less than the variation in lesion load.

A second study [62] by the MAGNIMS group into the much more subtle parameter MTR showed intercentre variation could be enormous compared with the relatively subtle effects caused by disease. The major sources of variation were the pulse sequence and MT pulse shape and size, and these varied considerably between MRI manufacturers. The EuroMT pulse sequence was produced [8] through the detailed study of the sources of variation. The intercentre variation could be reduced to relatively small amounts in the order of 3 pu using this sequence. The residual difference was attributed to B_1 error and non-uniformity. By correcting for B_1 variations the mean MTR value intercentre difference was made insignificant [103]. A recent detailed analysis and refinement of a procedure for measuring brain MTR histograms resulted in a method that gave identical histograms at two different sites with imagers from different manufacturers. This probably represents the ultimate example of how intercentre differences can be eliminated [12].

In any attempt to minimise intercentre differences, the initial effort should be focused on matching the sequence parameters and image analysis procedure as far as possible and this will be reasonably successful. However, it is likely that only by addressing B_1 effects, which vary from machine to machine (through RF coil design, pre-scan procedure, slice profile), can excellent multicentre agreement be achieved. Achieving good multicentre performance is the ultimate test of how good our quantitative techniques are. The early history of such attempts shows intercentre differences can often be large and with hindsight it is now understood that without attention to the issues described in this paper, intercentre effects are likely to be large enough to surprise many researchers.

We can expect multicentre studies to become easier, as researchers and manufacturers become more aware of the issues involved in good quantification. The recent MTR histogram study indicates a successful outcome using this approach [12], although the MRI physics effort

Table 5. 15 principles for multicentre studies

Initial design	
1	Measure quantitative parameters (such as T_1 , apparent diffusion coefficient or magnetisation transfer ratio value) rather than signal intensity
2	Understand the instrumental factors that will cause the measured values to vary (e.g. B_1 value, sequence timing, gradient strength)
3	Ideally use the same MRI manufacturer
4	Use the same RF coil configuration (ideally body-coil transmission with receive-only head coil)
5	Use sequences and sequence parameters that are similar
6	Be aware that in clinical trials centres are sometimes chosen by pharmaceutical companies either on the basis of collecting as many patients as possible for the trial or for fostering relations with clinicians who will later on be in a position to prescribe the pharmaceutical under test. Thus, a large number of centres may become involved, and the quality of the MRI data may become compromised by heterogeneity in the MRI centres. Selecting a subset of these, on the basis of MRI quality and homogeneity, may actually increase the power of the MRI study, as well as reduce the effort required to validate the centres and read their data
Initial evaluation of centres	
7	Validate by measuring control subjects, if appropriate, either individually at each centre, or by sending the same ones round to each centre. If the between-normal-subject variation is small, then different controls can be used at each centre, which is clearly convenient for an initial evaluation. If the normal variation is significant, then the same controls have to be sent round to each centre, which is logistically more complex and expensive. Ideally, at each institution, use the same radiographic (technical) staff
8	Consider imaging phantoms (test objects), which can either be made up for each centre, or sent round to each centre. If simple and easily reproduced, such as simple geometric objects or liquids, they can be provided at each centre, and provide an easy way to make an initial evaluation of the procedure. However, be aware that phantoms are not always stable and realistic
9	Record all relevant aspects of the data collection process, in particular those that are not automatically recorded in the image header file. (There has been a case of the wrong gadolinium dose being given in a multicentre study, which was only discovered by chance after the event because there was no record in the header file)
10	In the statistical analysis required for validation, measure separately the effects of the same observer repeating the analysis, of different observers carrying out the analysis, the same scanner repeating data collection, and different scanners carrying out data collection
11	If the interobserver effects are large, consider sending all the data (as computer files) to one centre for analysis by a single observer (or a convergent set of observers at one site). Sending a single dataset around to the different observers may clarify the source of variation
12	Be aware that sending data between sites requires expertise in reading image files (although the DICOM format is making this easier). The image headers should be read to ensure that the correct sequence parameters were used.
Analysis phase	
13	Standardise the creation of regions of interest, image segmentation, and all other aspects of the analysis. Create all regions of interest for each individual patient's serial trial data at one sitting [104]
14	Use automatic recording of the analysis procedure so that all aspects can be recreated
15	Use appropriate study design and statistical analysis techniques so that small residual intercentre effects can be accommodated without producing false positive results or reducing sensitivity

needed to achieve this was significant. Body coil excitation (used with multi-array receive coils) will give more uniform B_1 fields. At 3T and above, B_1 effects become important and although fast B_1 measurement techniques are now available [27] which enable residual B_1 effects to be measured and corrected, 1.5T operation is likely to be more reliable. Increased demand from quantification-aware users will (hopefully) drive the manufacturers to improve their products (we can already see that provision of spectroscopy and DTI acquisition has become routine).

A summary of issues to be aware of when undertaking a multicentre study is shown in Table 5. Provided the between-centre variation has been controlled (e.g. to <5%) then the remaining variation can probably be dealt with by using appropriate study design and statistics. The power of the study is then usually limited by the within-subject instrumental variation which is determined by repeated scans.

Acknowledgments

Dr Laura Parkes (University of Manchester, UK) kindly supplied the material on CBF measured using ASL. Dr Nicholas Dowell contributed material on diffusion QA, from a forthcoming book chapter on Diffusion QA [107]. Dr Dan Tozer and Mr Chris Benton assisted in the study on tumour volume using water bottle phantoms. Ms Rebecca Haynes and Professor Mara Cercignani kindly contributed literature.

References

1. Tofts PS. Standardisation and optimisation of magnetic resonance techniques for multicentre studies. *J Neurol Neurosurg Psychiatry* 1998;64:S37–43.
2. Tofts PS. QA: Quality assurance, accuracy, precision and phantoms. In: Tofts PS, ed. *Quantitative MRI of the brain: measuring changes caused by disease*. Chichester, UK: John Wiley, 2003:55–81.

3. Tofts PS. The measurement process: MR data collection and image analysis. In: Tofts PS, ed. *Quantitative MRI of the brain: measuring changes caused by disease*. Chichester, UK: John Wiley, 2003:17–54.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
5. Bland JM, Altman DG. Measurement error. *BMJ* 1996; 312:1654.
6. Grimaud J, Lai M, Thorpe J, Adeleine P, Wang L, Barker GJ, et al. Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magn Reson Imaging* 1996;14:495–505.
7. Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ, Miller DH. Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain* 1995;118:1601–12.
8. Barker GJ, Schreiber WG, Gass A, Ranjeva JP, Campi A, van Waesberghe JH, et al. A standardised method for measuring magnetisation transfer ratio on MR imagers from different manufacturers—the EuroMT sequence. *Magn Reson Mater Phy* 2005;18:76–80.
9. Filippi M, Gawne-Cain ML, Gasperini C, vanWaesberghe JH, Grimaud J, Barkhof F, et al. Effect of training and different measurement strategies on the reproducibility of brain MRI lesion load measurements in multiple sclerosis. *Neurology* 1998;50:238–44.
10. Filippi M, Horsfield MA, Ader HJ, Barkhof F, Bruzzi P, Evans A, et al. Guidelines for using quantitative measures of brain magnetic resonance imaging abnormalities in monitoring the treatment of multiple sclerosis. *Ann Neurol* 1998;43:499–506.
11. Jasperse B, Valsasina P, Neacsu V, Knol DL, De SN, Enzinger C, et al. Intercenter agreement of brain atrophy measurement in multiple sclerosis patients using manually-edited SIENA and SIENAX. *J Magn Reson Imaging* 2007;26:881–5.
12. Tofts PS, Steens SC, Cercignani M, Admiraal-Behloul F, Hofman PA, van Osch MJ, et al. Sources of variation in multicentre brain MTR histogram studies: body-coil transmission eliminates inter-centre differences. *MAGMA* 2006;19: 209–22.
13. d'Arcy JA, Collins DJ, Padhani AR, Walker-Samuel S, Suckling J, Leach MO. Informatics in Radiology (infoRAD): Magnetic Resonance Imaging Workbench: analysis and visualization of dynamic contrast-enhanced MR imaging data. *Radiographics* 2006;26:621–32.
14. Mulkern RV, Forbes P, Dewey K, Osganian S, Clark M, Wong S, et al. Establishment and results of a magnetic resonance quality assurance program for the pediatric brain tumor consortium. *Acad Radiol* 2008;15:1099–110.
15. Jackson JS, Tozer D, Tofts PS. Rapid measurement of subtle sub-one-percent changes in transmitter output and receive gain to monitor scanner stability. *Proceedings of the International Society Magnetic Resonance in Medicine*, 14th annual meeting, Seattle. 2006; 2363.
16. Jackson JS, Tozer D, Tofts PS. Measurement of subtle scanner instability using a stable thermally-insulated phantom. *Proceedings of the International Society Magnetic Resonance in Medicine*, 13th annual meeting, Miami. 2005; 2290.
17. Dowell NG, Tofts PS. Simple Reliable and Precise Quantitative Quality Assurance of in-vivo Brain ADC. *Proceedings of the International Society Magnetic Resonance in Medicine*, 16th annual meeting, Toronto. 2008; 3152.
18. Tofts PS, Lloyd D, Clark CA, Barker GJ, Parker GJ, McConville P, et al. Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo. *Magn Reson Med* 2000;43:368–74.
19. Leach MO, Brindle KM, Evelhoch JL, Griffiths JR, Horsman MR, Jackson A, et al. The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. *Br J Cancer* 2005;92:1599–610.
20. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging* 1999;10:223–32.
21. Galbraith SM, Lodge MA, Taylor NJ, Rustin GJ, Bentzen S, Stirling JJ, et al. Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. *NMR Biomed* 2002;15:132–42.
22. Buckley DL. Uncertainty in the analysis of tracer kinetics using dynamic contrast-enhanced T₁-weighted MRI. *Magn Reson Med* 2002;47:601–6.
23. Padhani AR, Hayes C, Landau S, Leach MO. Reproducibility of quantitative dynamic MRI of normal human tissues. *NMR Biomed* 2002;15:143–53.
24. Kostler H, Ritter C, Lipp M, Beer M, Hahn D, Sandstede J. Prebolus quantitative MR heart perfusion imaging. *Magn Reson Med* 2004;52:296–9.
25. Risse F, Semmler W, Kauczor HU, Fink C. Dual-bolus approach to quantitative measurement of pulmonary perfusion by contrast-enhanced MRI. *J Magn Reson Imaging* 2006;24:1284–90.
26. Parker GJ, Barker GJ, Tofts PS. Accurate multislice gradient echo T(1) measurement in the presence of non-ideal RF pulse shape and RF field nonuniformity. *Magn Reson Med* 2001;45:838–45.
27. Dowell NG, Tofts PS. Fast, accurate, and precise mapping of the RF field in vivo using the 180 degrees signal null. *Magn Reson Med* 2007;58:622–30.
28. Stevenson VL, Parker GJ, Barker GJ, Birnie K, Tofts PS, Miller DH, et al. Variations in T₁ and T₂ relaxation times of normal appearing white matter and lesions in multiple sclerosis. *J Neurol Sci* 2000;178:81–7.
29. Rutgers DR, van der GJ. Relaxation times of choline, creatine and N-acetyl aspartate in human cerebral white matter at 1.5 T. *NMR Biomed* 2002;15:215–21.
30. Ethofer T, Mader I, Seeger U, Helms G, Erb M, Grodd W, et al. Comparison of longitudinal metabolite relaxation times in different regions of the human brain at 1.5 and 3 Tesla. *Magn Reson Med* 2003;50:1296–301.
31. Deoni SC. High-resolution T₁ mapping of the brain at 3T with driven equilibrium single pulse observation of T₁ with high-speed incorporation of RF field inhomogeneities (DESPO T₁-HIFI). *J Magn Reson Imaging* 2007; 26:1106–11.
32. Deoni SC, Williams SC, Jezzard P, Suckling J, Murphy DG, Jones DK. Standardized structural magnetic resonance imaging in multicentre studies using quantitative T₁ and T₂ imaging at 1.5 T. *Neuroimage* 2008;40: 662–71.
33. Benner T, Reimer P, Erb G, Schuierer G, Heiland S, Fischer C, et al. Cerebral MR perfusion imaging: first clinical application of a 1 M gadolinium chelate (Gadovist 1.0) in a double-blinded randomized dose-finding study. *J Magn Reson Imaging* 2000;12:371–80.
34. Rana AK, Wardlaw JM, Armitage PA, Bastin ME. Apparent diffusion coefficient (ADC) measurements may be more reliable and reproducible than lesion volume on diffusion-weighted images from patients with acute ischaemic stroke-implications for study design. *Magn Reson Imaging* 2003;21:617–24.
35. Marks MP, de Crespigny A, Lentz D, Enzmann DR, Albers GW, Moseley ME. Acute and chronic stroke:

- navigated spin-echo diffusion-weighted MR imaging. *Radiology* 1996;199:403–8.
36. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. *Radiology* 1996;201:637–48.
 37. Filippi M, Iannucci G, Cercignani M, Assunta RM, Pratesi A, Comi G. A quantitative study of water diffusion in multiple sclerosis lesions and normal-appearing white matter using echo-planar imaging. *Arch Neurol* 2000;57:1017–21.
 38. Ciccarelli O, Werring DJ, Wheeler-Kingshott CA, Barker GJ, Parker GJ, Thompson AJ, et al. Investigation of MS normal-appearing brain using diffusion tensor MRI with clinical correlations. *Neurology* 2001;56:926–33.
 39. Cercignani M, Inglese M, Pagani E, Comi G, Filippi M. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. *AJNR Am J Neuroradiol* 2001;22:952–8.
 40. Zhang L, Harrison M, Heier LA, Zimmerman RD, Ravdin L, Lockshin M, et al. Diffusion changes in patients with systemic lupus erythematosus. *Magn Reson Imaging* 2007;25:399–405.
 41. Welsh RC, Rahbar H, Foerster B, Thurnher M, Sundgren PC. Brain diffusivity in patients with neuropsychiatric systemic lupus erythematosus with new acute neurological symptoms. *J Magn Reson Imaging* 2007;26:541–51.
 42. Emmer BJ, van der Grond J, Steup-Beekman GM, Huizinga TW, van Buchem MA. Selective involvement of the amygdala in systemic lupus erythematosus. *PLoS Med* 2006;3:e499.
 43. Clark CA, Hedehus M, Moseley ME. In vivo mapping of the fast and slow diffusion tensors in human brain. *Magn Reson Med* 2002;47:623–8.
 44. Steens SC, Admiraal-Behloul F, Schaap JA, Hoogenraad FG, Wheeler-Kingshott CA, Le Cessie S, et al. Reproducibility of brain ADC histograms. *Eur Radiol* 2004;14:425–30.
 45. Wheeler-Kingshott C, Barker GJ, Steens SCA, van Buchem MA. D: the diffusion of water. In: Tofts PS, ed. *Quantitative MRI of the brain: measuring changes caused by disease*. Chichester, UK: John Wiley, 2003: 203–56.
 46. Marengo S, Rawlings R, Rohde GK, Barnett AS, Honea RA, Pierpaoli C, et al. Regional distribution of measurement error in diffusion tensor imaging. *Psychiatry Res* 2006;147:69–78.
 47. Pfefferbaum A, Adalsteinsson E, Sullivan EV. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J Magn Reson Imaging* 2003;18:427–33.
 48. Holz M, Heil SR, Sacco A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate H-1 NMR PFG measurements. *Phys Chem Chem Phys* 2000;2:4740–2.
 49. Laubach HJ, Jakob PM, Loevblad KO, Baird AE, Bovo MP, Edelman RR, et al. A phantom for diffusion-weighted imaging of acute stroke. *J Magn Reson Imaging* 1998;8:1349–54.
 50. Delakis I, Moore EM, Leach MO, De Wilde JP. Developing a quality control protocol for diffusion imaging on a clinical MRI system. *Phys Med Biol* 2004;49:1409–22.
 51. Mills R. Self-diffusion in normal and heavy water in the range 1–45°. *J Phys Chem* 1973;77:685–8.
 52. Tofts PS, Jackson JS, Tozer DJ, Cercignani M, Keir G, MacManus DG, et al. Imaging cadavers: cold FLAIR and noninvasive brain thermometry using CSF diffusion. *Magn Reson Med* 2008;59:190–5.
 53. Rashid W, Hadjiprocopis A, Davies G, Griffin C, Chard D, Tiberio M, et al. Longitudinal evaluation of clinically early relapsing-remitting multiple sclerosis with diffusion tensor imaging. *J Neurol* 2008;255:390–7.
 54. Bisdas S, Bohning DE, Besenski N, Nicholas JS, Rumboldt Z. Reproducibility, interrater agreement, and age-related changes of fractional anisotropy measures at 3T in healthy subjects: effect of the applied b-value. *AJNR Am J Neuroradiol* 2008;29:1128–33.
 55. Oouchi H, Yamada K, Sakai K, Kizu O, Kubota T, Ito H, et al. Diffusion anisotropy measurement of brain white matter is affected by voxel size: underestimation occurs in areas with crossing fibers. *AJNR Am J Neuroradiol* 2007;28:1102–6.
 56. Chang LC, Jones DK, Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. *Magn Reson Med* 2005;53:1088–95.
 57. Ozturk A, Sasson AD, Farrell JA, Landman BA, da Motta AC, Aralasmak A, et al. Regional differences in diffusion tensor imaging measurements: assessment of intrarater and interrater variability. *AJNR Am J Neuroradiol* 2008;29:1124–7.
 58. Yanasak NE, Allison JD, Hu TC. An empirical characterization of the quality of DTI data and the efficacy of dyadic sorting. *Magn Reson Imaging* 2008;26:122–32.
 59. Fieremans E, De Deene Y, Delputte S, Ozdemir MS, D'Asseler Y, Vlassenbroeck J, et al. Simulation and experimental verification of the diffusion in an anisotropic fiber phantom. *J Magn Reson* 2008;190:189–99.
 60. Sasaki M, Yamada K, Watanabe Y, Matsui M, Ida M, Fujiwara S, et al. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. *Radiology* 2008;249:624–30.
 61. Cercignani M, Bammer R, Sormani MP, Fazekas F, Filippi M. Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. *AJNR Am J Neuroradiol* 2003;24:638–43.
 62. Berry I, Barker GJ, Barkhof F, Campi A, Dousset V, Franconi JM, et al. A multicenter measurement of magnetization transfer ratio in normal white matter. *J Magn Reson Imaging* 1999;9:441–6.
 63. Silver NC, Barker GJ, MacManus DG, Tofts PS, Miller DH. Magnetisation transfer ratio of normal brain white matter: a normative database spanning four decades of life. *J Neurol Neurosurg Psychiatry* 1997;62:223–8.
 64. Davies GR, Altmann DR, Hadjiprocopis A, Rashid W, Chard DT, Griffin CM, et al. Increasing normal-appearing grey and white matter magnetisation transfer ratio abnormality in early relapsing-remitting multiple sclerosis. *J Neurol* 2005;252:1037–44.
 65. Mendelson DA, Heinsbergen JF, Kennedy SD, Szczepaniak LS, Lester CC, Bryant RG. Comparison of agarose and cross-linked protein gels as magnetic resonance imaging phantoms. *Magn Reson Imaging* 1991;9:975–8.
 66. Tozer D, Ramani A, Barker GJ, Davies GR, Miller DH, Tofts PS. Quantitative magnetization transfer mapping of bound protons in multiple sclerosis. *Magn Reson Med* 2003;50:83–91.
 67. Tofts PS, Cercignani M, Tozer DJ, Symms MR, Davies GR, Ramani A, et al. Quantitative magnetization transfer mapping of bound protons in multiple sclerosis. *Magn Reson Med* 2003;50:83–91. Erratum in: *Magn Reson Med* 2005;53:492–3.
 68. Tozer D, Benton CE, Cercignani M, Tofts PS, Rees J. Quantitative Magnetisation Transfer Parameters show Dramatic Changes in Low Grade Gliomas. *Proceedings of the International Society Magnetic Resonance in Medicine*, 15th annual meeting, Berlin. 2007; 2852.
 69. McKnight TR. Proton magnetic resonance spectroscopic evaluation of brain tumor metabolism. *Semin Oncol* 2004; 31:605–17.

70. Payne GS, Leach MO. Applications of magnetic resonance spectroscopy in radiotherapy treatment planning. *Br J Radiol* 2006;79:S16–26.
71. Kreis R. Quantitative localized 1H MR spectroscopy for clinical use. *Prog Nucl Magn Reson Spectrosc* 1997;31:155–95.
72. Tofts PS, Waldman AD. Spectroscopy: ¹H metabolite concentrations. In: Tofts PS, ed. *Quantitative MRI of the brain: measuring changes caused by disease*. Chichester, UK: John Wiley 2003:299–339.
73. Leach MO, Collins DJ, Keevil S, Rowland I, Smith MA, Henriksen O, et al. Quality assessment in in vivo NMR spectroscopy: III. Clinical test objects: design, construction, and solutions. *Magn Reson Imaging* 1995;13:131–7.
74. van der Graaf M, Julia-Sape M, Howe FA, Ziegler A, Majos C, Moreno-Torres A, et al. MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. *NMR Biomed* 2008;21:148–58.
75. Keevil SF, Barbiroli B, Collins DJ, Danielsen ER, Hennig J, Henriksen O, et al. Quality assessment in in vivo NMR spectroscopy: IV. A multicentre trial of test objects and protocols for performance assessment in clinical NMR spectroscopy. *Magn Reson Imaging* 1995;13:139–57.
76. Garcia-Gomez JM, Luts J, Julia-Sape M, Krooshof P, Tortajada S, Robledo JV, et al. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *MAGMA* 2009;22:5–18.
77. Soher BJ, Hurd RE, Sailasuta N, Barker PB. Quantitation of automated single-voxel proton MRS using cerebral water as an internal reference. *Magn Reson Med* 1996;36:335–9.
78. Keevil SF, Barbiroli B, Brooks JC, Cady EB, Canese R, Carlier P, et al. Absolute metabolite quantification by in vivo NMR spectroscopy: II. A multicentre trial of protocols for in vivo localised proton studies of human brain. *Magn Reson Imaging* 1998;16:1093–106.
79. Brooks WM, Friedman SD, Stidley CA. Reproducibility of 1H-MRS in vivo. *Magn Reson Med* 1999;41:193–7.
80. Schirmer T, Auer DP. On the reliability of quantitative clinical magnetic resonance spectroscopy of the human brain. *NMR Biomed* 2000;13:28–36.
81. Naressi A, Couturier C, Castang I, de Beer R, Graveron-Demilly D. Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. *Comput Biol Med* 2001;31:269–86.
82. Provencher SW. Automatic quantitation of localized in vivo 1H spectra with LCModel. *NMR Biomed* 2001;14:260–4.
83. Jager HR, Waldman AD, Benton C, Fox N, Rees J. Differential chemosensitivity of tumor components in a malignant oligodendroglioma: assessment with diffusion-weighted, perfusion-weighted, and serial volumetric MR imaging. *AJNR Am J Neuroradiol* 2005;26:274–8.
84. Rees J, Watt H, Jager HR, Benton C, Tozer D, Tofts P, et al. Volumes and growth rates of untreated adult low-grade gliomas indicate risk of early malignant transformation. *Eur J Radiol* 2009;72:54–64.
85. Galanis E, Buckner JC, Maurer MJ, Sykora R, Castillo R, Ballman KV, et al. Validation of neuroradiologic response assessment in gliomas: measurement by RECIST, two-dimensional, computer-assisted tumor area, and computer-assisted tumor volume methods. *Neuro Oncol* 2006;8:156–65.
86. Shah GD, Kesari S, Xu R, Batchelor TT, O'Neill AM, Hochberg FH, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neuro Oncol* 2006;8:38–46.
87. Bynevelt M, Britton J, Seymour H, MacSweeney E, Thomas N, Sandhu K. FLAIR imaging in the follow-up of low-grade gliomas: time to dispense with the dual-echo? *Neuroradiology* 2001;43:129–33.
88. Tofts PS, Barker GJ, Filippi M, Gawne-Cain M, Lai M. An oblique cylinder contrast-adjusted (OCCA) phantom to measure the accuracy of MRI brain lesion volume estimation schemes in multiple sclerosis. *Magn Reson Imaging* 1997;15:183–92.
89. Parkes LM. Quantification of cerebral perfusion using arterial spin labeling: two-compartment models. *J Magn Reson Imaging* 2005;22:732–6.
90. Xie J, Gallichan D, Gunn RN, Jezzard P. Optimal design of pulsed arterial spin labeling MRI experiments. *Magn Reson Med* 2008;59:826–34.
91. Alsop DC, Detre JA. Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow. *J Cereb Blood Flow Metab* 1996;16:1236–49.
92. Luh WM, Wong EC, Bandettini PA, Hyde JS. QUIPSS II with thin-slice T1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling. *Magn Reson Med* 1999;41:1246–54.
93. Ye FQ, Mattay VS, Jezzard P, Frank JA, Weinberger DR, McLaughlin AC. Correction for vascular artifacts in cerebral blood flow values measured by using arterial spin tagging techniques. *Magn Reson Med* 1997;37:226–35.
94. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magn Reson Med* 1998;40:383–96.
95. Parkes LM, Tofts PS. Improved accuracy of human cerebral blood perfusion measurements using arterial spin labeling: accounting for capillary water permeability. *Magn Reson Med* 2002;48:27–41.
96. 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;51:736–43.
97. Petersen ED, Golay X. Is Arterial Spin Labeling Ready for Primetime? Preliminary Results from the QUASAR reproducibility study. *Proceedings of the International Society Magnetic Resonance in Medicine*, 16th annual meeting, Toronto 2008; 191.
98. Miles KA. Perfusion imaging with computed tomography: brain and beyond. *Eur Radiol* 2006;16:S37–43.
99. Hoeffner EG, Case I, Jain R, Gujar SK, Shah GV, Deveikis JP, et al. Cerebral perfusion CT: technique and clinical applications. *Radiology* 2004;231:632–44.
100. Szucs-Farkas Z, Kurmann L, Strautz T, Patak MA, Vock P, Schindera ST. Patient exposure and image quality of low-dose pulmonary computed tomography angiography: comparison of 100- and 80-kVp protocols. *Invest Radiol* 2008;43:871–6.
101. Tan JS, Tan KL, Lee JC, Wan CM, Leong JL, Chan LL. Comparison of Eye Lens Dose on Neuroimaging Protocols between 16- and 64-Section Multidetector CT: Achieving the Lowest Possible Dose. *AJNR Am J Neuroradiol* 2009;30:373–7.
102. Goh V, Halligan S, Bartram CI. Quantitative tumor perfusion assessment with multidetector CT: are measurements from two commercial software packages interchangeable? *Radiology* 2007;242:777–82.
103. Ropele S, Filippi M, Valsasina P, Korteweg T, Barkhof F, Tofts PS, et al. Assessment and correction of B(1) – induced errors in magnetization transfer ratio measurements. *Magn Reson Med* 2005;53:134–40.

104. Beresford MJ, Padhani AR, Taylor NJ, Ah-See ML, Stirling JJ, Makris A, et al. Inter- and intraobserver variability in the evaluation of dynamic breast cancer MRI. *J Magn Reson Imaging* 2006;24:1316–25.
105. O'Connor JPB, Tofts PS, Miles KA, Parkes LM, Thompson G and Jackson A. Dynamic contrast-enhanced imaging techniques: CT and MRI. *Br J Radiol* 2011;84(Spec. No. 2):S112–20.
106. Reynolds G, Wilson M, Peet A, Arvanitis TN. An algorithm for the automated quantitation of metabolites in *in vivo* NMR signals. *Magn Reson Med* 2006;56: 1211–19.
107. Dowell NG, Tofts PS. Quality assurance for diffusion MRI. In Jones DK, ed. *Diffusion: theory, methods and applications*. New York: Oxford University Press, 2011: 319–30.