2

The Measurement Process: Magnetic Resonance Data Collection and Image Analysis

AQ: Please check that "MR" is correctly spelled out as "Magnetic Resonance" (throughout).

Contents

Paul S. Tofts
Brighton and Sussex Medical School

2.1 Magnetic Resonance Data Collection

The process of collecting magnetic resonance (MR) data from a subject, in the form of images, spectra or maps, is analysed in some detail.

2.1.1 Subject Positioning and the Prescan Procedure

The subject is positioned on the scanner couch by the radiographer (technologist). The subject should be comfortable, to reduce movement during the scan as much as possible. The radiographer should use any insight into the subject's emotional state to reduce anxiety if necessary; preparation on a separate couch may be helpful. A cushion under the knees can reduce cramp. Occasionally it is desirable to place the subject prone. *Prone positioning* of the head may be more comfortable if support is provided for the forehead and cheekbones, leaving a gap for the nose, as used in a massage table. *Movement* of the body can cause a head movement; a nasal positioning device (Tofts *et al.*, 1990) can help cooperative subjects to keep still. Some patients will find it hard to keep still because of their disease; researchers involved in the study are usually motivated to keep very still. Some kinds of movement are very common, especially rotation in the sagittal plane ('nodding'). Movement can be monitored by repeated localiser images throughout the study. If spatial registration between different image datasets is used (see Chapter 17), then the amount of movement that took place is available as output from the program. Research on both why some subjects move and on what limits how long a subject can stay in the scanner would improve the quality of MR data that can be obtained. It may become possible to use fast MR (or optical) imaging to dynamically alter the slice positions, tracking the movement of a subject in real time (although movement to a location of different static or RF field value would require some sophisticated correction). If Gd contrast agent is to be injected, a line is placed into the subject's arm, so that injection is carried out without disturbing the positioning of the subject. A power injector is usually used to provide a consistent injection procedure, with synchronisation to the scanner.

After the subject has been placed in the magnet bore, the automatic *prescan* procedure generally includes the following steps, which take account of differences between subjects and are crucial to quantification. The receiver gain is adjusted to use the available dynamic range of the receiver channel, without overloading it. The gain must be fixed for subsequent scans, if image intensity values are to be combined in some way (e.g. for a dynamic Gd scan series, where images are collected at a range of time points after injection of contrast agent – see Chapter 14). The transmitter output is adjusted to give the desired flip angle (FA) in the subject. This can be carried out in a number of ways; ideally, only the signal from the relevant piece of tissue (e.g. a slice or a spectroscopic voxel) is optimised. A multislice or volume acquisition cannot have the correct FA at all locations, because of transmit field non-uniformity.

The pulse sequences, containing long lists of radio frequency (RF) and magnetic field gradient pulses, are then run. Signals are recorded; localisation of the origin of signal is achieved using a combination of slice or slab selection, frequency encoding and phase encoding gradients. Images can be weighted by various parameters (e.g. T_1, T_2 or D – see the chapters on each MR parameter). Images are reconstructed using Fourier transformation; the magnitude of the complex data is usually calculated (this is not vulnerable to unpredictable phase shifts). Full descriptions of the MR imaging process are available elsewhere (Brown *et al.*, 2014, and also see Table in Chapter 1).

AQ: Please check "see Table in Chapter 1" for clarity; please provide table number

2.1.2 The NMR Signal

The signal δv from precessing nuclei in a small volume δV_s in the sample is given by the following (Hoult and Richards 1976; Hoult 1978):

$$\delta \upsilon = \omega_0 B_{1xy} M_{xy} \delta V_s \cos(\omega_0 t) \tag{2.1}$$

where ω_0 is the Larmor[1] frequency (in radians s^{-1}); the life of Sir Joseph Larmor, the Irish physicist, is described by Tubridy and McKinstry (2000). B_{1xy} is the component[2] of the RF field B_1 produced in the transverse plane at the location of the sample by unit current in the coil, during transmission.[3] M_{xy} is the transverse component of the magnetization of the sample.[4] For protons the equilibrium magnetization M_0 is as follows (Brown *et al.*, 2014):

$$M_0 = \frac{N \gamma h^2 B_0}{4kT} \tag{2.2}$$

where N is the number of protons per unit volume, γ is the *magnetogyric ratio*,[5] h_bar [printer - h-bar] = h/2π, where h is Planck's constant, B_0 is the magnitude of the main static magnetic field, k is Boltzmann's constant and T is the absolute temperature[6] of the sample.

AQ: Please check "where h is Planck's constant" for clarity; please suggest whether "where" may be deleted.

The proportionality of received signal with the magnitude of the applied field per unit current, shown in Equation 2.1, is called the *principle of reciprocity* (Hoult and Richards 1976), and this has been a key concept in quantitative MR. In simple terms, it says that if we have trouble getting the applied B_1 field into a particular location in the sample, using a particular coil, we will have as much trouble getting the signal out of that location, using the same coil. This is discussed in more detail in Section 2.1.10 below.

AQ: Please check whether the sentence "if we have trouble getting the applied B_1 field into a particular location" is okay as edited.

The dependence of magnetization on absolute temperature is relevant when room temperature concentration standards are used (as in measurements of proton density and metabolite concentrations). As a particular concentration of protons is cooled (for example from body temperature to room temperature), its magnetization increases and it can produce more signal (see Chapter 3).

2.1.3 The Static Magnetic Field B_0

In a superconducting magnet, the value of the static field is set at the time of installation by adjusting the amount of circulating current stored in the windings. There may be a very small decay over time, which is compensated for by adjusting the current through room temperature windings or by adjusting the centre frequency of the transmitter.

When the subject is placed in the magnet, the magnetic susceptibility of the tissue alters the field inside the brain slightly. The transmitter centre frequency is adjusted to bring the protons back onto resonance. The shim coil currents are adjusted to obtain a spatially uniform B_0 distribution, as far as possible.

Remaining static field gradients caused by spatially varying tissue susceptibility (particularly near tissue–air interfaces, such as the temporal lobes) can be a problem, particularly for spectroscopy and echo planar imaging, which are very sensitive to such gradients. In spectroscopy the line position will be altered and possibly broadened. In gradient echo and echo planar imaging there may be signal dropout due to intravoxel dephasing.[7] In spin echoes, the dephasing effect of these gradients is corrected provided the spins are stationary; however, in the presence of diffusion, spins moving through a gradient will not be rephrased and signal loss will once again be seen. Such signal loss will not normally cause systematic error in quantification, although the lowered signal-to-noise ratio (SNR) will give increased random errors, and in situations where the absolute signal level is important (e.g. PD) there will also be a systematic error.

AQ: Please spell out PD at first mention

A further source of degradation is that echo planar images (and to a lesser extent gradient echo images, which have a much shorter echo time than echo planar sequences) will suffer geometric distortion, such that the image is shifted or warped in the locality of susceptibility gradients (Moerland *et al.*, 1995; Hutton *et al.*, 2002; Jezzard 2002). This in turn prevents straightforward spatial registration of such images with those having negligible

[1] The *Larmor frequency* is the frequency at which protons precess around the main static field B_0.

[2] A linear coil produces two counter-rotating components; one is in the right direction for NMR and is useful; the other is not used but contributes to noise and power requirements. In a circularly polarised coil only the useful component is produced and detected.

[3] This equation is valid for the simple case of a single transmit/receive coil at low field. In this case the transmit field B_1^+ equals the receive field B_1^- (i.e. $B_1^+ = B_1^- = B_{1xy}$); see also Section 2.1.10.

[4] After a single 90° RF pulse, $M_{xy} = M_0$.

[5] $\gamma = \omega_0/B_0$, where B_0 is the static magnetic field strength, in Tesla. For protons, $\gamma = 2.675\ 10^8$ rad s^{-1} T^{-1} (equivalent to 42.57 MHz/T) (Brown *et al.* 2014). Greek letters are described in Appendix 1.

[6] The *absolute temperature* is measured in degrees Kelvin (K) from –273°C, which is called *absolute zero*. Thus the freezing point (0°C) is 273K, and body temperature (37°C) is 310K.

[7] In *intravoxel dephasing*, the different components of magnetization in a voxel, experiencing different static fields, become out of phase with each other, and the total transverse magnetization vector in the voxel is reduced. In a spin echo, this dephasing is corrected by the 180° refocusing pulse; in a gradient echo the uncorrected dephasing leads to signal loss.

distortion (principally those that are spin-echo–based, although gradient echo sequences often also have negligible distortion) and thwarts any attempt at measuring volume. The image intensity is likely to be altered by distortion (since a given amount of signal will be placed into a voxel that is too large or too small). A third degradation is that off-resonance effects in such localities may reduce the apparent FA and distort a 2D slice selection process (see Section 2.1.6).

The static field can be mapped straightforwardly using the phase shift after a gradient echo (Sled and Pike 2000; Hetherington *et al.*, 2006).

2.1.4 Static Field Gradients

Having taken a lot of care to achieve a uniform static magnetic field, switched field gradients[8] are deliberately introduced as part of the imaging process. The slew rates are very fast, giving typical switching times of <100 μs. Eddy currents can be induced in surrounding conducting structures; these have the effect of producing small transient shifts in B_0, distorting spectra and images. Eddy currents are reduced to low levels by several devices. Actively shielded gradient coils limit the magnetic flux outside the coil; current pre-emphasis circuits drive the coils in such a way as to counteract the effects of the eddy currents; and conduction loops are eliminated from the scanner bore construction materials.

The remaining non-idealities are twofold. Firstly, the *gradient amplitudes* may be incorrect, by up to about 1% (depending on the calibration procedure). This gives rise to small errors in the size of objects (since the gradient change corresponds to a change in magnification, or of voxel size) and also to errors in the estimates of diffusion coefficients and tensors. Very precise measurements of voxel dimensions, using image registration, enable the value of small gradient changes to be measured (Lemieux and Barker 1998) Secondly, gradient coils do not produce a completely linear variation of static field with distance (i.e. the gradients are non-uniform); this in turn produces errors in the gradient amplitude (according to the position) and gives rise to spatial distortion (Moerland *et al.*, 1995; Jezzard 2002). *Non-linearity from gradient coils* is minimal in the central (head) region when body gradient coils. Manufacturers usually measure and correct for the geometric distortion caused by non-linearity; this can be seen by turning the correction off.

AQ: Please check the sentence 'Non-linearity from gradient coils is minimal in the central (head) region when body gradient coils' for clarity

2.1.5 Radio Frequency Transmit Field B_1^+

Usually transmission is by a body coil, which has relatively good uniformity over the head region. Typical pulse amplitudes are 10–20 μT. Sometimes amplitudes are expressed in hz or radians s^{-1}, giving the rate of nutation of magnetization around a constant RF field of that value.[9] The current in the transmit coil required to achieve this value of RF field depends on the Q[10] of the coil. Q is determined mostly by power losses in the subject, caused mostly by its electrical conductivity rather than by losses in the coil itself. As the subject is moved into the coil, its conductivity loads the coil, and a greater current, and hence voltage, is required to produce a given B_1 value. The amount of *coil loading* (i.e. the amount of power that is removed from the coil and is deposited in the subject) varies from subject to subject. The prescan procedure sets out to obtain the same value of B_1 (and hence FA), regardless of the loading produced by the particular subject. The *transmitter output* is adjusted, usually automatically, often by adjusting *attenuators* in the amplifier.

Non-linearity in the transmitter output stage may occur, leading to incorrect B_1 values and distortion of selective pulses. Such gross non-linearity over the normal range of amplitudes of the selective pulses would probably be picked up as artefacts in the routine imaging (depending on the amplitudes of the selective pulses used). Calibration of the transmitter output stage is usually carried out periodically as part of routine preventive maintenance (Venkatesan *et al.*, 1998). *Transmitter linearity* can be investigated as follows: An oscilloscope can be used to measure the output voltage as a function of software hard pulse amplitude (Alecci *et al.*, 2001); this will be accurate to within a few percent, depending on the oscilloscope. The output can be stepped by the software for convenience. A more accurate method is to use NMR to measure the B_1 amplitude, as follows. At each amplitude, observe the signal from a small sample as a function of hard pulse duration. The null duration (i.e. a 180° pulse) gives an accurate and precise measurement of B_{1+}. Ensure there is no pulse droop, using an oscilloscope. A plot of B_1^+ vs software pulse amplitude should be linear.

RF non-uniformity is the largest cause of error in qMR.

> Radiofrequency field inhomogeneities are the most irksome sources of nonidealities (especially as a result of their omnipresence). The spatial variation of the RF field sensitivity of the transmit and receive coils enter the signal expression for any sequence in the form of altering the flip angle at a given spatial location as well as altering the received signal from [the same] spatial location. (Haacke *et al.*, 1999 p661)

At 1.5T the effect is noticeable (Barker *et al.*, 1998); at higher fields the problem becomes worse (see Figure 2.1). An elliptical object (such as the head) in a circularly polarised coil gives a diagonal non-uniformity pattern (Sled and Pike 1998). At 3T using a birdcage coil a 20% reduction in B_1^+ was measured at the

[8] Imaging gradients can be up to about 20 mT m^{-1}; diffusion gradients are often higher (to shorten the echo time), up to 80 mT m^{-1}, and may use a dedicated head coil set. Switching (slew) rates are up to 200 T m^{-1} s^{-1}.

[9] The nutation rate is $\omega_1 = \gamma B_1$ rad s^{-1}; thus a RF field of 10 μT corresponds to 2680 rad s^{-1} or 426 hz. The duration of a hard (i.e. non-selective) θ pulse is $\tau_\theta = \theta/(\gamma B_1)$; thus a 90° 10 μT hard pulse lasts 587 us.

[10] *Q* stands for *quality factor*, denoting how long a coil will ring after being excited. High-Q coils are less damped, lose less energy per cycle, ring for longer, provide a greater B_1 for a given current and provide a greater signal for a given amount of precessing magnetisation.

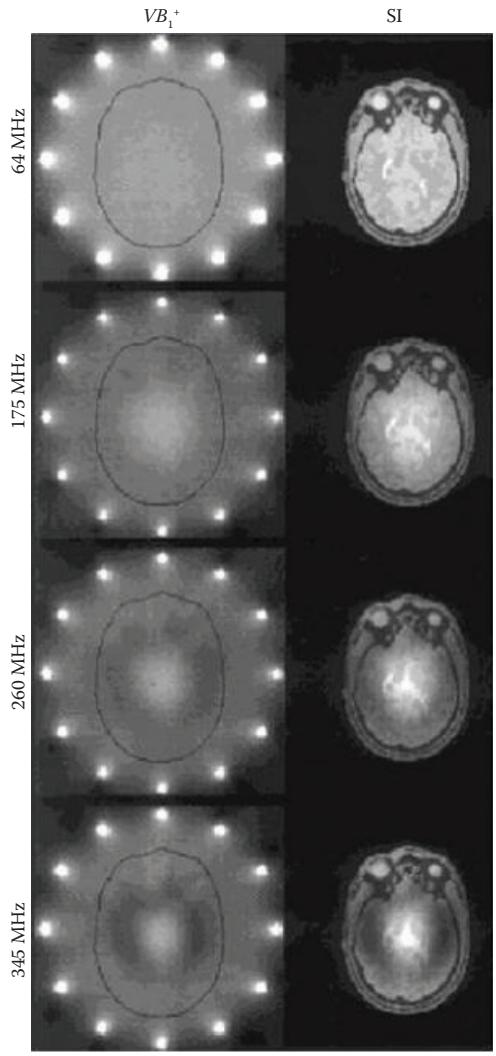


FIGURE 2.1 RF signal non-uniformity in the head. An accurate mathematical model of the head inside a birdcage coil, with flip angle = 90° at the head centre. Frequencies correspond to fields of 1.5T, 4.1T, 6.1T and 8.1T. VB_1^+ is the excitation field (normalised by the factor V), SI is the signal intensity from a gradient echo sequence. The doming effect at the centre of the head becomes increasingly pronounced at higher fields, and an annular region of reduced signal is visible further from the centre. (From Collins, C.M., and Smith, M.B., *Magn. Reson. Med.*, 45(4), 684–691, 2001.)

periphery of the brain, compared to its value at the centre (Alecci *et al.*, 2001). Anatomically accurate models of B_1^+ distribution in the head, using detailed anatomical knowledge (Collins and Smith 2001; Ibrahim *et al.*, 2001) show that at high fields, up to 8T, non-uniformity increases, as dielectric resonance increases the sensitivity near the centre of the head; measurements at 7T confirm this (Collins *et al.*, 2002). Adjustment of the current applied through each port on a multiple transmit coil, and its phase, allows uniformity to be optimised (Ibrahim *et al.*, 2001). This has been termed *RF shimming* (Collins *et al.*, 2005).

In the early days of MRI, it was feared that the reduced RF penetration at high frequencies (the 'skin effect') (Bottomley and Andrew 1978) caused by the electrical conductivity of the tissue would prevent head imaging above 20 MHz. In fact it is more than offset by the amplifying effect of dielectric resonance (which increases B_1^+ in objects whose size is comparable with the half-wavelength of electromagnetic waves at the frequency of observation; see Figure 2.1).

2.1.6 Slice and Slab Profile

In 2D (slice selective) imaging, slice selection is a key problem. The observed transverse magnetization is the sum of spins within the voxel that have experienced a variety of histories, according to their location in the slice selective gradient, the local FA and the amount of relaxation (i.e. T_1/TR). Thus the slice profile can be distorted; there is not a single effective FA within a voxel, and signal modelling for quantification is often complex and inaccurate (Parker *et al.*, 2001).[11] Slice selection is often used for an EPI readout and also is needed for some MR parameters (e.g. T_2, sT_2^*); then such effects must be taken into account. Magnetization preparation by a hard (non-selective) pulse followed by 2D readout can mitigate the distortion.

AQ: Please spell out TR at first mention

AQ: Please spell out EPI at first mention

Fast and powerful gradients have driven the widespread use of 3D imaging, which has no such problems. Slab selection (in which signal from outside the desired field of view is suppressed) is usually applied in each phase encoding direction, to prevent wrap-around.[12]

2.1.7 B_1+ Transmit Field Mapping

The determination of the RF 'active' field B_1+ at each location in space is important for two reasons. First, the local FA can then be found; this is needed for many parameter calculations (e.g. T_1 from VFA Chapter 5[13]). Second, its value is needed for accurate measurement of some MR parameters (e.g. MT).

AQ: Please spell out MT

Early methods of determining FA, or setting it to a required value, were developed in spectroscopy. A sample inside a uniform RF coil gives a maximum signal when the FA is 90°, provided it can relax fully between pulses (i.e. TR >> T1). The pulse amplitude,[14] or duration, is increased from a low value; at first

[11] Slice distortion and correction is described and illustrated in more detail in the 1st edition.

[12] Slab selection is a more gentle process than slice selection (each voxel has a single well-defined FA) and probably produces little distortion; however, there seems to be no literature on this.

[13] Chapter 5 also contains a discussion of B_1^+ mapping.

[14] A rectangular ('hard') pulse, of amplitude B_1 and duration τ, produces a FA of $\gamma B_1 \tau$. Depending on the spectrometer hardware, either the pulse amplitude or duration is varied to give the required FA. Early spectrometers had fixed amplitude, and the duration was altered. Modern MRI machines often also allow variable amplitude (since a selective ['soft'] pulse must keep its duration fixed).

AQ: Please spell out TR, EPI-DA beneath Table 2.1

AQ: Please provide missing column heads for Table 2.1

TABLE 2.1 B_1^+ Mapping Methods (A Selection)

Magnitude Methods		
Actual flip imaging	Yarnykh 2007	Two pulses, same FA, two TRs (3D)
	Hurley *et al.*, 2012	Combined with T_1 measurement
Dual angle method (DAM)	Stollberger and Wach 1996	Two pulses, two FAs, long TR (2D)
	Insko and Bolinger 1993	Original
	Cunningham *et al.*, 2006	Saturate for short TR
	Boudreau *et al.*, 2017	EPI-DA; fast uses epi
180° null	Dowell and Tofts 2007	Can use short TR
Phase Methods		
Bloch-Siegert	Sacolick *et al.*, 2010	Phase α B_1^2. 3D
	Sacolick *et al.*, 2011	3 sec acquisition for prescan
Phase sensitive	Morrell 2008	Wider range than DAM

FA, flip angle.

the signal increases almost linearly, then reaches a maximum, then declines. Further increase in the pulse gives a null signal (corresponding to an FA of 180°), and this condition can often be found more precisely than the maximum at 90°.

Modern methods (summarised in Table 2.1) should meet the following nine criteria: the method should (i) provide an accurate value, to within about 1% (since some measurements are very sensitive to FA errors, e.g. T_1[15]); (ii) have good precision;[16] (iii) have reasonable imaging time (ideally less than 1 minute; this favours 3D methods); (iv) be independent of T_1 effects; (v) be independent of B_0 (off-resonance) effects; (vi) be unaffected by 2D slice selection artefacts; (vii) have acceptable SAR value; (viii) work over a wide FA range if needed;[17] and (ix) ideally be capable of implementation using a standard pulse sequence.

AQ: Please spell out SAR at first mention

Early work often used the double-angle method (Stollberger and Wach 1996). Two 2D acquisitions are made, with nominal FA values typically 30° and 60°; the ratio of signals gives the actual FA values but only under the condition of TR >> T1 (complete relaxation). The needs for speed and 3D acquisition may have pushed this method aside. However, an EPI variant provides a 2-minute solution using a standard pulse sequence, and a performance comparable to the AFI and BS methods (Boudreau *et al.*, 2017) (Figure 2.2).

AQ: Please confirm whether the inserted figure citations for the figures are okay.

The two principle methods currently in use are actual flip angle imaging (AFI) and Bloch–Siegert (BS) (Sacolick *et al.*, 2010 2011; Whisenant *et al.*, 2016). AFI (Yarnykh 2007) uses two pulses with the same FA, at different (short) TRs, in a 3D sequence. There are residual T_1 effects, which could be measured using VFA (variable flip angle – see Chapter 5). In VAFI (Hurley *et al.*, 2012), VFA and AFI are combined to take proper account of T_1. AFI can be seen as an ingenious variant on the dual angle method. The ratio of signals from two acquisitions is again used; however, it is the TR not the FA that is altered. In the BS approach, an off-resonant pulse is used to produce a phase that is proportional to the square of the local B_1^+ field; this is then read out. It is fast and independent of B_0 error (Sacolick *et al.*, 2010 2011; Duan *et al.*, 2013; Whisenant *et al.*, 2016).

AQ: Please check that DAM has been correctly spelled out as 'dual angle method'

The 180° null method uses a standard 3D sequence to provide a map in 4 minutes – see Figure 2.3 (Dowell and Tofts 2007). The move to higher fields and the consequent increase in B_1^+ non-uniformity has prompted MRI manufacturers to invest in mapping techniques (Sacolick *et al.*, 2010; Nehrke and Börnert 2012). Summaries with discussion, comparisons and often optimisation of sequences are given by Lutti *et al.*, (2010), Morrell and Schabel (2010), Sacolick *et al.*, (2010), Volz *et al.*, (2010), Hurley *et al.*, (2012), Park *et al.*, (2013) and Pohmann and Scheffler (2013).[18]

Inflow effects: In the special case of bulk blood flow, the effective B_1^+ value may be smaller than that in the voxel of interest. ASL[19] uses this effect; blood arrives in the voxel with a different magnetization from local water that has a recent history in the voxel. In DCE imaging,[20] blood is often imaged with a 2D slice to measure its T_1; however, blood from outside the slice may have a lower FA and hence higher magnetization than that predicted by a naive use of the FA value in the voxel.[21]

AQ: Please spell out ASL at first mention

AQ: Please spell out DCE at first mention

2.1.8 B_1^- Receive Sensitivity Field

Reception is usually by a close-fitting multiple array head coil, for good sensitivity. The receive sensitivity field $B_1^-(r)$ is determined by both the receive coil array geometry and the properties of the object being imaged (whether a head or a phantom). Dielectric resonance tends to give a high B_1^- in the central parts of the object (similar to the B_1^+ distribution), whilst multiple surface coils have greater response from the periphery; these effects can be made to partially cancel (Figure 2.4).

In a multi-array receive system, parallel imaging options (Grappa, Sense, Smash, etc.) can produce image artefacts, arising from the reconstruction of low spatial frequencies, which might be unstable and degrade quantification performance. Thus parallel imaging should be used with caution, with low acceleration factors.

B_1^- cannot be determined directly and with the loss of the reciprocity principle[22] has become one of MRI's 'unsolved problems'. For most MR parameters it is a ratio of signals that is measured,

[15] A 1% error in FA gives a 2% error in T_1; see Chapter 5, Equation 5.10.

[16] The B_1 function will vary slowly with space, so some spatial smoothing is permissible.

[17] Although some methods presume a small range of FA values, they can usually be adapted to a wider range by adding more measurement pulses.

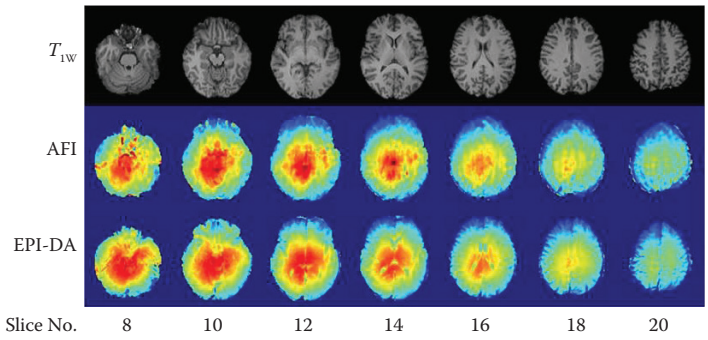
[18] Teaching sessions at the annual International Society for Magnetic Resonance in Medicine meeting by PF van de Moortele (2015) and Lawrence L Wald (2016) were also insightful.

[19] See Chapter 16.

[20] See Chapter 14, Section 14.4.

[21] Blood flowing at 1 m s^{-1} into a 5 mm 2D slice will spend only 5 ms in the slice; thus it may experience only one pulse (if TR = 5 ms), insufficient to reach a new equilibrium magnetization.

[22] See Section 2.1.10.



AQ: Please spell out EPI-DA

AQ: Boudrea et al. (2017): Please provide volume number and page range.

FIGURE 2.2 B_1^+ maps at 3T using the actual flip angle imaging and EPI-DA methods (see Table 2.1) (From Boudreau, M., *et al.*, *J. Magn. Reson. Imaging*, 2017).

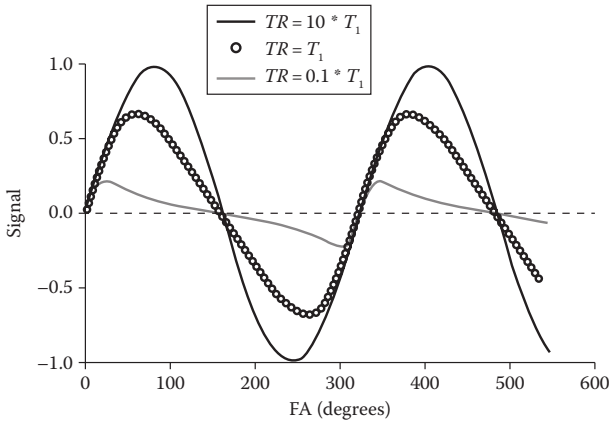


FIGURE 2.3 Simulation showing how the 180° signal null can be used to map B_1, regardless of T_1. (From Dowell, N.G., and Tofts, P.S., *Magn. Reson. Med.*, 58(3), 622–630, 2007.)

and absolute sensitivity has no influence. However when measuring absolute concentrations of protons (i.e. PD and magnetic resonance spectroscopy (MRS) absolute metabolite concentrations), a map of $B_1^-(r)$ is required, and some attempts have been made to estimate it.

The *bias field* approach uses information that $B_1^-(r)$ varies slowly with position (Volz *et al.*, 2012; see also Watanabe *et al.*, 2011; Jin *et al.*, 2012; Sabati and Maudsley 2013). A PD-weighted image has intensity proportional to $B_1^-(r)PD(r)$; smoothing this (fwhm = 60 mm) gives a map proportional to $B_1^-(r)$ (this assumes PD(r) has little low spatial frequency content). A single reference value for PD (e.g. from cerebrospinal fluid or a mean value for the whole brain) then enables $B_1^-(r)$ to be determined and its effect removed in a PD(r) estimation. The method relies on PD(r) being smooth and there being no large abnormalities. A variation on this is to constrain the possible behaviour of $B_1^-(r)$ by using the Maxwell equations; the method has been successful in the ordered environment of a single transmit/receive coil and a phantom (Sbrizzi *et al.*, 2014).

The *receiver gain* can be altered during the prescan procedure to account for the magnitude of the signal (the gain is much reduced in spectroscopy). Ideally it will be fixed during the acquisition of the image series,[23] at a suitable value that will not overload the receive chain nor introduce extra noise. If it is altered, a correction might be possible, depending on what information on receiver gain is available and whether analogue values of attenuation or gain are accurate.

2.1.9 Image Noise

Electrical noise comes from random thermal agitation (Brownian motion) in the subject, the RF coil and possibly the preamplifier. With good design, contributions from the hardware are made insignificant and the dominant source is the subject.

Artefacts (often from movement during the phase encoding process) constitute an additional source of unpredictable error; these can be assessed by viewing the air surrounding the head. The centre of the grey level display window is set to zero. Any visible artefacts are of interest, since they almost certainly extend to the high SNR parts of the image, in the brain, and may exceed the random noise.

2.1.9.1 Optimised Sequence Parameters

The effect of image noise can be reduced by careful choice of sequence parameters. Increasing the voxel size reduces noise (at the expense of spatial resolution). Increasing the number of averages (NEX – the number of excitations) reduces the noise (SNR α $NEX^{1/2}$). Increasing the TR often also increases the SNR. Changing NEX or TR usually increases the acquisition time, and for a fixed acquisition time an optimal combination of parameters can be found, which maximises SNR. The acquisition of an image dataset from which parameters such as blood–brain permeability are to be estimated can also be optimised (Tofts 1996) (see Figure 2.5 and Chapter 3 section on modelling error).

AQ: Please check the phrase "section on modelling error" for clarity; please provide section number here for reference.

[23] For example, during acquisition of several TE values to determine T_2 (Chapter 6).

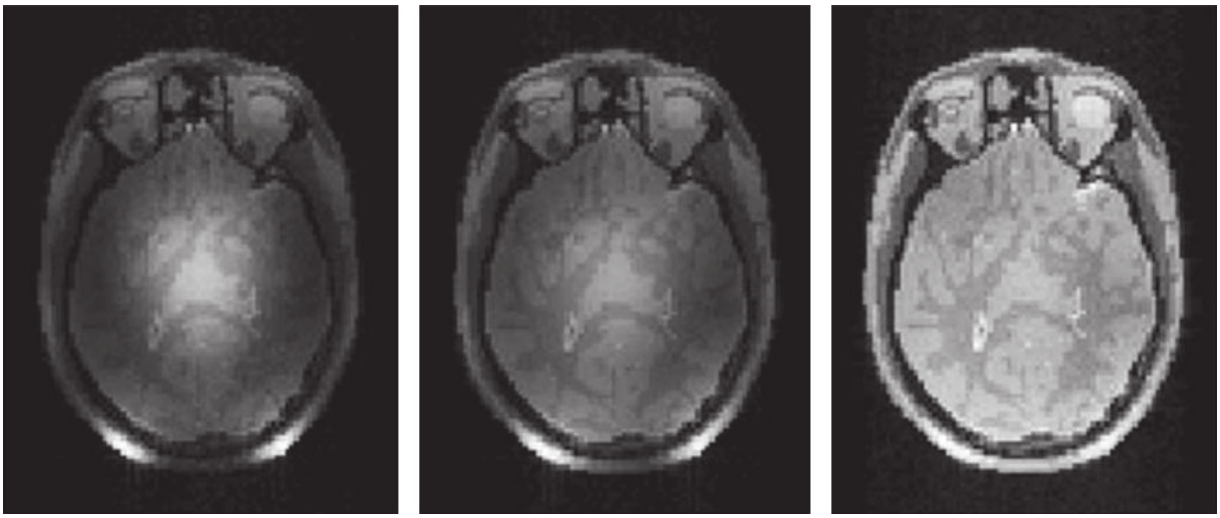


FIGURE 2.4 Simulated magnetic resonance images showing the results of different coil configurations at 300 MHz (7T). Use of a single volume coil in both transmission and reception (a) results in relatively strong B_1^+ and B_1^+ fields near the centre (due to constructive interference) surrounded by weaker fields (due to destructive interference) resulting in a centre-bright appearance of the signal intensity distribution. Use of a single volume coil in transmission but an array of decoupled coils in reception with sum-of-magnitude reconstruction (b) results in relatively strong B_1^+ near the centre but relatively strong B_1^- near the periphery, resulting in a more homogeneous signal intensity distribution. Use of a transmit array with RF shimming to lessen the pattern of constructive and destructive interference in transmission and a receive array with sum-of-magnitude reconstruction (c) produces a very homogeneous image, even at this high frequency with only eight elements in transmission and reception. (From Collins, C.M., and Wang, Z., *Magn. Reson. Med.*, 65(5), 1470–1482, 2011.)

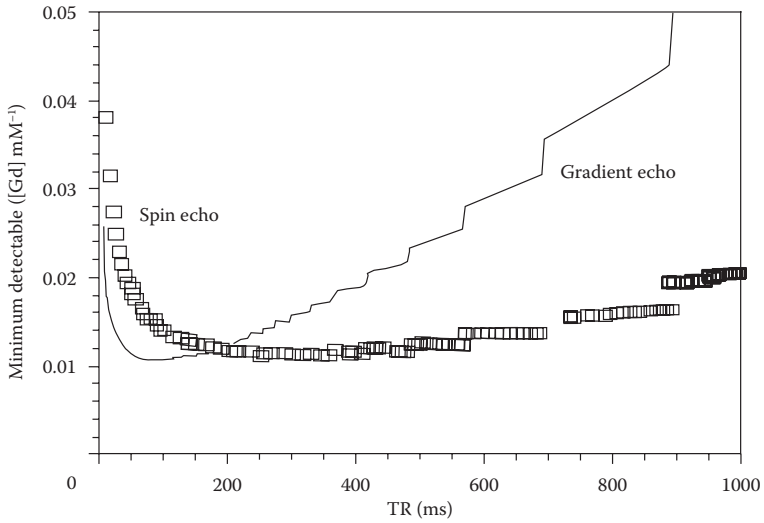


FIGURE 2.5 Sequence optimisation by noise modelling. Mathematical modelling of image noise propagation predicts the minimum amount of Gd contrast agent that can be detected using a T_1-weighted sequence. By optimizing the repetition time, *TR*, in a spin echo or gradient echo sequence, its performance in white matter can be optimised. Theory indicates that, for a spin echo, the optimum $TR = T_1/2$ (here it was assumed $T_1 = 600$ ms). The gradient echo (flip angle, FA = 50°) can achieve the same sensitivity, provided the correct *TR* is used. The examination time was fixed at 10 min. (From Tofts, P.S., *Magn. Reson. Imaging*, 14(4), 373–380, 1996.)

2.1.9.2 Rician Noise Distribution in Magnitude Image Gives Systematic Error

Most images are constructed from the magnitude of the complex image data, and phase information is discarded. Magnitude data does not have a normal distribution (it cannot be negative). At low values of SNR, the distribution from a single receive coil follows a Rician probability distribution with a non-zero mean value (Henkelman 1985, 1986) (Brown *et al.*, 2014). At high values of SNR this approximates to a Gaussian (normal) distribution; at zero SNR it becomes a *Rayleigh distribution* (Figure 2.6).

This effect constitutes a systematic error, or bias, in addition to the random error that noise always contributes. It can be seen most clearly by looking at the mean value in air regions of the image, where a non-zero value will be found. Techniques that use low SNR image data, principally T_2 and ADC, will be affected. The applications where this is most important are in T_2 and diffusion measurements, where the decay of signal may be followed down into the noise (Miller and Joseph 1993; Wheeler-Kingshott *et al.*, 2002), and perfusion, where the signal difference is comparable with the noise (Karlsen *et al.*, 1999). If image averaging is used to improve SNR, this should be carried out on the complex images, before formation of the magnitude.

Correcting single-coil image data is possible; however in multicoil receive systems correction is not straightforward and additionally the noise may vary with position in the image (see Chapter 17).

2.1.9.3 Noise Estimation

An estimate of the noise value in high SNR regions is often desirable (e.g. for modelling error propagation). For a single coil system, if the standard deviation in high SNR regions of the image is σ, then sampling the Rician noise in an air region of interest (ROI) will give a mean value of 1.25σ and a standard deviation of 0.66σ (Edelstein *et al.*, 1984; Gudbjartsson and Patz 1995; Andersen 1996), enabling σ to be estimated. Noise has also been estimated by subtracting the squares of images (Sijbers *et al.*, 1998).

In a multicoil receive system, two approaches are possible. (1) Placing an ROI in a uniform region of tissue will give an indication of image noise; however the SD value will also contain a contribution from any tissue non-uniformity in the ROI, and the multiple coil system will in any case cause the noise value to vary with position. (2) Subtraction of two repeated images is likely to give a better result. The difference image is expected to have a normal distribution with an SD of $\sqrt{2}\,\sigma$; the distribution of values, and their mean, should be checked for any unexpected effects. This method has been used in phantoms (Murphy *et al.*, 1993; Goerner and Clarke 2011). The noise in brain images is likely to be different (because of different coil loading and B_1^- distribution), and an explicit measurement in brain is preferred. Provided movement is controlled, and measurements are restricted to uniform areas of tissue, then this ought to be possible.

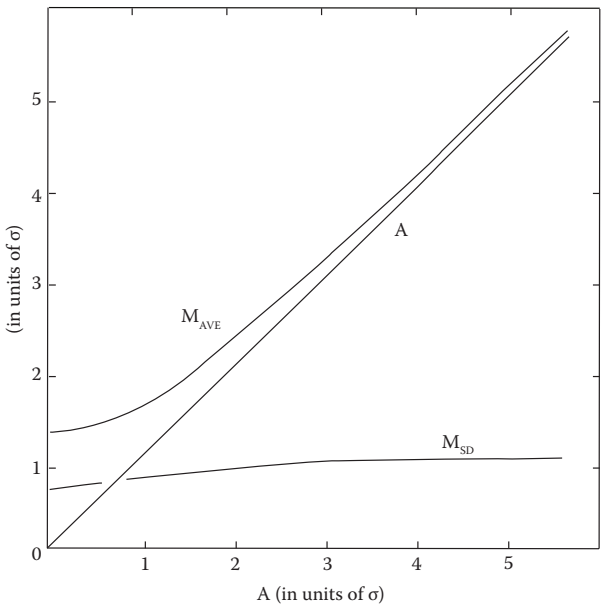


FIGURE 2.6 Rician noise in single-coil magnitude data. For a uniform region of an image with real amplitude A (measured in units of σ, the standard deviation of the noise in each dimension of the normal distribution), the average value M_{AVE} and the standard deviation M_{SD} in the magnitude (M) image are shown, both expressed in units of σ. Thus at high signal-to-noise (SNR) ratio ($A >> \sigma$), the mean intensity in the magnitude image equals its value in a real image ($M_{AVE} = A$), whilst at low SNR it exceeds the value in a real image, reaching an asymptotic value of 1.253σ in the absence of any signal. At high SNR, the standard deviation equals that in a real image ($M_{SD} = \sigma$), whilst in the absence of any signal it decreases to $M_{SD} = 0.655\sigma$. (From Henkelman, R.M., *Med. Phys.*, 12(2), 232–233, 1985.)

2.1.9.4 Image Quantization Errors

Image intensity values are usually stored as integers, typically to 12-bit precision (i.e. 1 in 4096). The floating point values from the image reconstruction process are rounded to the nearest integer.[24] Typical signal values may be 500–1000; thus the electrical noise will be a few units (for SNR = 100), and quantization noise (maximum value 0.5) should be insignificant compared to electrical noise. If the floating point numbers are truncated (not rounded) a small amount of bias (0.5 image units) will be introduced.

If histograms are to be produced from parameter maps, *image despiking*[25] should be carried out, to prevent the discrete image probability distribution from producing artefacts in the histograms (Tozer and Tofts 2003).

2.1.10 The Reciprocity Principle and Its Failure

In the early days of MRI, the reciprocity theorem was key in enabling B_1^- (r) to be found. (Hoult and Richards 1976; Hoult 1978, 2000). In this case $B_1^- = B_1^+ = B_1$, and in this paradigm many papers described 'B_1 mapping' (when in fact they are mapping B_1^+). Reciprocity was demonstrated at fields up to 1.5T (Tofts and Wray 1988; Michaelis *et al.*, 1993; Barker *et al.*, 1998; Fernandez-Seara *et al.*, 2001). Under varying values of coil loading, the product of pulse length[26] or transmitter output voltage and signal is constant. Given B_1^-, absolute measurements of PD and MRS metabolite concentration were possible (Provencher 1993, 2001; Fernandez-Seara *et al.*, 2001).[27] Reciprocity served us well for 40 years.

Later it became clear that the theorem is not valid at higher fields (3T and above),[28] i.e. $B_1^- \neq B_1^+$ (Sled and Pike 1998; Hoult 2000; Ibrahim 2005; Collins and Wang 2011), and also that as transmit and receive coils improved, single transmit/receive coils would go out of use. Now in the 'brave new technoworld' of higher fields and separate transmit and receive coils, the reciprocity theorem is probably consigned to the past, and we have to travel on without knowledge of B_1^- (Figure 2.7).

PD and MRS measurements have become more difficult with the loss of the principle of reciprocity. It is almost impossible to distinguish whether a change in signal voltage is caused by a change in proton concentration or a change in coil sensitivity. The solution may lie in using a dedicated 'reciprocity-friendly scanner' that works at 1.5T or less, with a single transmit/receive coil. Such a PD map could be matched to a higher quality proton-density weighted image from another scanner to provide improved spatial resolution and SNR.

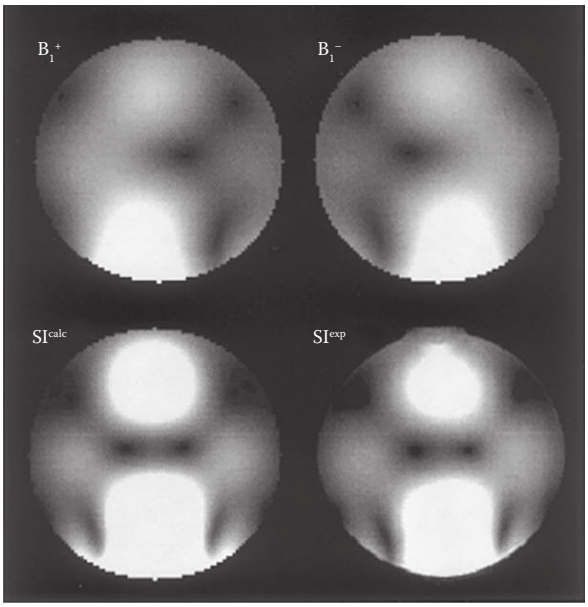


FIGURE 2.7 Reciprocity failure at 7T. Theoretical B_1 distributions in a 16 cm spherical NaCl solution phantom at 7T, imaged with a 10 cm surface coil. B_1^+ is the circularly polarised transmit field, B_1^- is the receive field. SI^{calc} and SI^{exp} are the theoretical (calculated) and measured (experimental) signal intensities for a gradient echo image; their similarity gives confidence in the modelling. (From Collins, C.M., *et al.*, *Magn. Reson. Med.*, 47(5), 1026–1028, 2002.)

2.1.11 Non-uniformity Correction

Early workers found that images were often visibly non-uniform, and much effort was devoted to measuring and correcting non-uniformity.[29] Image non-uniformity (NU) arises from three sources: (1) Transmit field (B_1^+) non-uniformity gives a magnetization M_{xy} NU that also depends on B_1 (unless a fully relaxed sequence is used). Receive (B_1^-) NU depends on both (2) the receive coil characteristics and on (3) the head electromagnetic characteristics. Thus a simple image of a uniform object, or simple smoothing, cannot correct for these three factors in a quantitative way. *Transmit non-uniformity* is usually minimised by using the body coil, and measured if necessary (see Section 2.1.7).

2.1.12 Scanner Stability

Despite having followed all good practice, image data can be unstable for unknown reasons. Repeated imaging of a phantom may show a short-term variation in signal that is greater than that predicted by image noise (Weisskoff 1996), in addition to long-term drift (see Figure 2.8). Such instability can have two major effects: (i) it degrades reproducibility (i.e. instrumental SD

AQ: Please check that "see Figure 2.8" is correct as edited; " [instab]" deleted.

[24] Rounding introduces a maximum error of 0.5, with an rms value of $1/\sqrt{120.} \approx 3$.

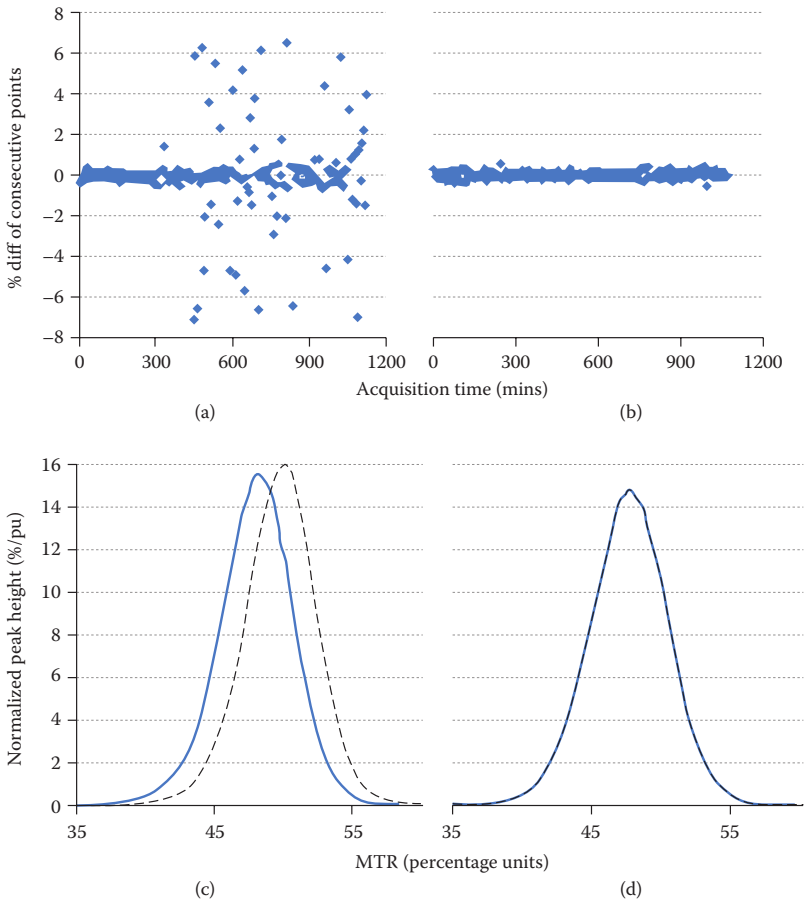
[25] The integers are converted to floating point numbers, and random noise with maximum magnitude 0.5 is added, forcing the image intensity values to have a continuous distribution, before calculation of the maps.

[26] Data from Tofts and Wray 1988, reanalyzed in the first edition, p 304.

[27] The 1st edition of this book contains much on reciprocity (p 37) and its application in PD (p95) and MRS (p 304–5).

[28] As the wavelength of the B_1 excitation decreases, the approximation of a quasistatic excitation field becomes less valid (from Sled and Pike, 1998). At 3T, the wavelength in water is 260 mm (see Chapter 3, Section 3.5.2).

[29] See 1st edition, p 37.



AQ: Please check whether the sentence 'Unsuspected scanner instability – an invisible problem. MTR histograms showed' is okay as edited.

FIGURE 2.8 Unsuspected scanner instability – an invisible problem. MTR histograms showed a large within-subject variation (c). Repeated scanning of a phantom overnight showed large random variation (a). After changing transmitter boards, the scanner was stable (b) and MTR histograms were reproducible (d). (Data from NG Dowell, originally presented in (Haynes *et al.*, 2010) (From Haynes, B.I.,*et al.*, Measuring scan-rescan reliability in quantitative brain imaging reveals instability in an apparently healthy imager and improves statistical power in a clinical study. ISMRM annual scientific meeting, Stockholm, p. 2999, 2010.)

AQ: Please check that ISD is correctly spelled out as "instrumental SD".

[ISD] – see Chapter 3, Section 3.3.2.1), and (ii) in a DCE series acquisition, it introduces time-dependent variation, which may mask a subtle signal enhancement caused by a genuine T_1 change. Thus if a parameter (e.g. T_1) shows an unexpectedly large[30] within-subject variation, it is worth carrying out repeated imaging and looking at the variation of the raw image data. If transmitter instability is suspected, then imaging at the *Ernst angle*[31] ought to produce a stable signal. If receiver instability is suspected, then slowly varying changes would affect the signals at all FAs equally.

AQ: Please check that "where" is correct as added to Footnote †.

2.2 Image Analysis, Statistics and Classification

2.2.1 Types of Image Analysis

Here an introduction to image analysis concepts is given; a higher-level viewpoint is given in Chapter 17. There are three main ways of extracting relevant image intensities from a set of images that may cover many slices, several tissue parameters and many subjects. Before analysis, the image dataset may be spatially registered. Images from a single subject may be registered, to reduce the effects of movement during the examination. Images from different subjects may be registered to a standard space.[32] Registration can produce subtle changes in image inten-

AQ: Please check whether hyphen may be removed from "MNI-space" in Footnote ‡.

[30] The effect of image noise σ on repeated measurements of an ROI mean is to give variation with a standard deviation equal to the standard error of the mean of the ROI (i.e. sem = σ/ √[no. of pixels]).

[31] In a spoilt gradient echo, the signal is maximal and independent of FA at $\cos(\theta_E) = \exp(-TR/T_1)$, where θ_E is the Ernst angle; e.g. for T_1 = 800 ms, TR = 10 ms, then $\theta_E = 9^o$.

[32] The most common is 'MNI-space' (MNI = Montreal Neurological Institute).

sity that may thwart attempts at quantification (e.g. in a DCE series, small changes in intensity can sometimes be observed that are caused by the registration process, not by changes in T_1).

2.2.1.1 Region of Interest Analysis

The study is focussed on a particular part or parts of the brain, for example visible lesions or large volumes of normal-appearing white matter, where intensities are to be measured. One or several ROIs are drawn for each subject. Regions can be circular, oval, square, rectangular or irregular. Regions may be defined in a single slice or extend over several slices (then the set is a volume of interest [VOI]). They are often created using a semi-automatic technique, which speeds up the process and improves the reproducibility. ROI size is a compromise between reducing noise (which favours large ROIs) and reducing partial volume error (which favours small ROIs). Alternatively, if the image datasets are in standard (stereotactic) space, then a number of standard VOIs will be available. The process of creating the ROIs takes some time to learn, and different observers will develop different approaches. Inter-observer[33] variation can be reduced by carefully defining the procedure to be used. This may include factors such as how the image is to be displayed, and a detailed description of what anatomical cues are to be used in positioning the ROI. Usually the intra-observer variation is lower, and many studies accept that a single observer should be used to analyse the whole dataset. Even for a single observer, the analysis should be repeated after a few days to ensure the reproducibility is reasonable. A formal measurement of reproducibility can be undertaken (see Chapter 3, Section 3.3).

Unbiased ROI generation: There may be multiple MR images or maps, for example conventional MRI (showing lesions in PD- or T_2-weighted images) and a MTR map. The appearance of lesions may be different on the conventional MRI and the map. If the MTR values of lesions are to be measured and tested, the ROIs should be defined on the conventional MR images (after spatial registration), then transferred to the maps. If the ROIs were defined directly on the maps, the map intensity would influence where the ROI boundary was placed. ROIs tend to be attracted to locations of abnormal intensity (as a result of the process of their creation, where an observer tends to draw around distinct objects). Thus any conclusions about map values in the lesions would be biased, since these values had been used to define which pixels would be included in the region. If serial measurements are made, fixed ROIs should be used for each time point, if conclusions are to be drawn about changes in a parameter value over time.[34] Large regions of most of the normal-appearing white or grey matter have been generated using T_1 maps (Parkes and Tofts 2002) or fractional diffusion anisotropy (Cercignani *et al.*, 2001).

AQ: Please spell out MTR at first mention

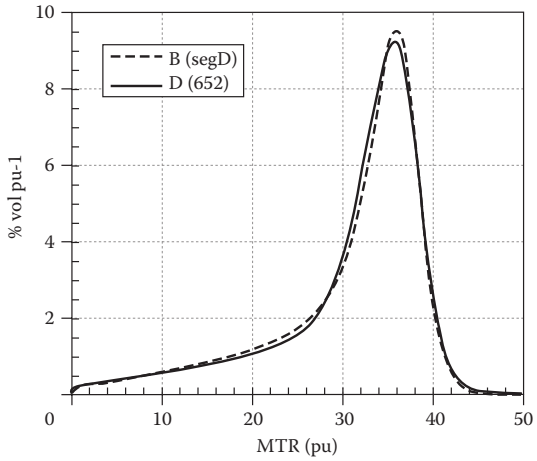


FIGURE 2.9 Matching MTR group histograms from two centres with 1.5T scanners from different manufacturers. By using body coil excitation and standardised histogram generation, inter-centre differences were eliminated. (From Tofts, P.S., *et al.*, *Magma*, 19(4), 209–222, 2006.)

In studies of diffuse disease that affects large parts of the brain, instead of creating large regions to study this, two other approaches are available: histograms and voxel-based group mapping.

2.2.1.2 Histogram Analysis

A solution to the problem of ROI placement, and possible bias arising from this process, is to test all of a tissue type (e.g. white matter). This is particularly appropriate for diseases where the biological effects are diffuse and widespread. Histograms do have the disadvantage that localisation information has been lost, and if disease only affects part of the region sensitivity will be reduced by pooling data from the whole region.

Histograms from different centres have sometimes varied; by standardising their generation, multicentre studies are possible (Tofts *et al.*, 2006). The bin width should be chosen to be small enough to capture any fine structure, yet large enough to not display statistical fluctuations caused by a small number of pixels in the bin (typically 5 ms for T_1). The bin should be labelled by its centre value (not the left- or right-hand edge). A normalised bin amplitude should be calculated as follows: find the percentage of the total pixels in that bin, then divide by the bin width. The total area under the histogram curve is then 100%, regardless of bin width (Figure 2.9).[35]

Histogram analysis by PCA and linear discriminant analysis can be very powerful; MTR histograms have predicted clinical score and also separated disease subtypes (Dehmeshki *et al.*, 2001 2002b) – see Figure 2.10.

AQ: Please spell out PCA at first mention

[33] *Inter-observer* means 'between observer', i.e. the difference between measurements made by *different* observers on the same image data. *Intra-observer* means 'within observer', i.e. the difference if the *same* observer repeats the measurement.

[34] A recent example showed that, after treatment of tumours, regions of Gd enhancement unexpectedly retained the same intensity. On being probed, the author disclosed that the ROIs had been drawn and redrawn on the Gd images and had become smaller after treatment. Thus fixed ROIs would have shown a reduction in enhancement.

[35] Examples are shown in the 1st edition, Chapter 18, along with examples of classification of clinical data.

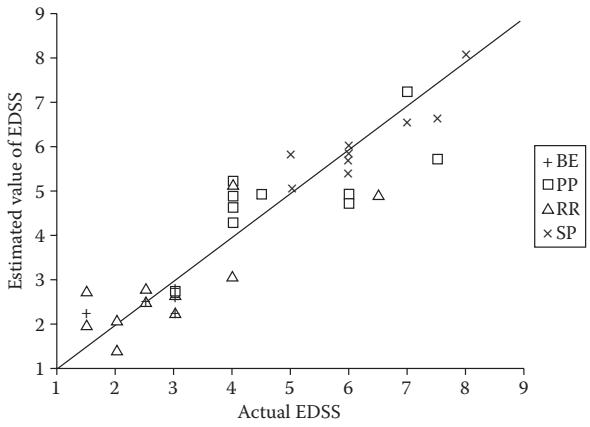


FIGURE 2.10 Correlation of Expanded Disability Status Scale (EDSS, a clinical score used in multiple sclerosis, MS) with principle components enables a 'predicted EDSS' to be calculated for each subgroup of people with MS, solely from the MTR histogram and knowledge of the subgroup. These are usually within one point of the actual clinically measured EDSS value. Compared with conventional features, PCA gives better correlation coefficients. (From Dehmeshki, J., *et al*., *Magn. Reson. Med*., 46(3), 600–609, 2001.)

AQ: Please spell out PCA at first mention.

2.2.1.3 Voxel-Based Group Mapping – Beyond ROIs and Histograms

Analysis of group-mapped images provides a way of combining the ability of ROIs to be spatially specific (and thus sensitive) with the ability of histograms to be unbiased. In essence, a complete set of ROIs is automatically generated at all locations across the whole brain, without any bias in where they are located. The image datasets are first spatially normalised to all lie in the same space. Appropriate statistical tests are then carried out on all the ROIs. Further information is in Chapter 17.

2.2.2 Types of Statistical Analysis

Statistical analysis of the measurements is an increasingly complex procedure. Involvement with statisticians at an early stage in study design and analysis is advisable. Failure to be aware of statistical pitfalls can lead to embarrassing rejections by journal referees; sometimes re-analysis can address the problems, sometimes data collection is fatally flawed. Clinical datasets can be very expensive to acquire (both financially and in terms of the effort by the patients and controls); this drives an imperative to make the best possible use of clinical data by using appropriate analysis techniques (see Chapter 1, Table 1.2).

In this section the basic concepts relevant to statistical analysis of MR data are summarised; however for full treatment the reader should go to the books (see Table 2.2). Suitable software packages are SPSS (originally 'Statistical Package for the Social Sciences' but now more general), SAS and STATA. These have comprehensive manuals associated with them, and the software is often available through academic sources. With such packages, 'clickity-click' analysis can be a hazard; a few mouse clicks can unleash a sophisticated analysis of a large dataset that the user does not understand and may misinterpret.

2.2.2.1 Group Comparisons – T-Test

The simplest test of a MR parameter is to compare measurements in two groups (typically diseased and 'normal'). Many MR parameters may have been measured. To test their usefulness, it is tempting to carry out multiple t-tests, for example to see if any of the parameters differ between clinical groups. If 20 tests are carried out at the $p = 0.05$ significance level, on average one test will come out positive by chance (this is known as a *type I error*). Thus the results of multiple comparisons must be treated with caution. The *Bonferroni correction* for multiple comparisons (Bland and Altman 1995) allow for this by suggesting that the appropriate p-value is the value that would have been used for a single test (e.g. $p = 0.05$), divided by the number of comparisons (e.g. 20). A much-reduced p-value (e.g. 0.002) is then used, and the chances of a type I error are reduced (in this example to 0.04). The Bonferroni correction can be unnecessarily cautious, missing an effect that is present (i.e. a *type II error*). If the outcome variables being tested are correlated, then the reduction of the p-value by the number of tests is too extreme. Conversely, if several of the tests show significance (which does imply correlation between the tested variables), then the chances of this occurring by chance are much lower than the chance of just one occurring by chance.

AQ: Please check the phrase "chances of this occurring by chance" for clarity; suggest e.g. "odds of this occurring by chance"

A useful distinction can be made between two kinds of study. A *fishing expedition* looks at many parameters, tests them at $p = 0.05$, accepting that some type I errors will occur, and uses this to gain insight or guide further studies. A strict *hypothesis-driven study* sets up the hypothesis *before-analysis*, makes only one test and is thus able to control type I and type II errors better. Thus a fishing expedition might be used to set up a hypothesis-driven study, which must be carried out on *separate data*. Alternatively, the data could be divided into two parts, the first used for fishing and the second for a strict test (although the power of the study would be reduced by the smaller sample sizes).

Negative results in a group comparison: A negative result has two possible explanations: firstly, the genuine biological spread in each group may be too large to pick out a significant different between them; secondly, the effect of measurement error may have broadened the spread in the groups, beyond its genuine biological value, enough to obscure genuine biological difference (see Chapter 3, Figure [group_overlap]), a kind of 'false-negative'. Thus a negative result can be the consequence of poor measurement technique (high ISD – see Chapter 3, Section 3.3.2.1). Another centre (with better technique) might succeed in showing a group difference in a similar group of subjects. The original centre might succeed with a larger sample (i.e. the study was underpowered). Failure to observe a difference does not mean that none exists.

AQ: Please replace the text "Figure [group_overlap]" with correct figure number

In the case of a negative result, insight can be obtained in two ways: first, estimates of measurement error and within-group variance should be made (see Chapter 3, Section 3.3.3). This enables the intraclass correlation coefficient (ICC) to be calculated; a good ICC means that the groups are genuinely indistinguishable;

AQ: Please provide heading for Table 2.2

TABLE 2.2

Title	Authors	Date Published	Number of Pages	Description
Statistical Methods in Medical Research	Peter Armitage, Geoffrey Berry, JNS Matthews,	2001	832	Classic from statisticians, 4th edition, hardback.
Practical Statistics for Medical Research	Douglas Altman	2000	254	2nd edition, paperback.
An Introduction to Medical Statistics	Martin Bland	2015	448	4th edition, paperback.
Essential Medical Statistics	B Kirkwood, J Sterne	2003	512	
Medical Statistics: A Textbook for the Health Sciences	MJ Campbell, DJ Machin, SJ Walters	2007	344	4th edition, paperback.
Health Measurement Scales: A Practical Guide to Their Development and Use	DL. Streiner, GR. Norman, J Cairney	2014	416	5th edition, good on clinical scales, paperback.
Analyzing Multivariate Data	J Lattin, JD Carroll, PE Green	2006	556	High level hardback, covers principal components analysis, analysis of variance, clustering, discriminant analysis and much more.
An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements	JR Taylor	1997	488	A clear description of eternal truths.

a poor ICC means that the failure to distinguish may be caused by poor instrumentation. Second, the confidence limits on the group means, and the minimum detectable group difference, should be reported. Other workers can then judge whether improved technique (i.e. reduced measurement error) might enable them to obtain a positive result.

Positive results: If a positive result is obtained, the confidence limits on group means and group difference should be given; then other groups can estimate whether their measurement errors are low enough to repeat the positive observation. False positive results can be obtained if there is another (confounding) factor that differs between the groups. This could be the time of scanning (if one group is scanned before a change in measurement procedure, and the other after the change) or other uncontrolled variables, e.g. age, gender, lifestyle differences or even head size.[36]

2.2.2.2 Correlation with Clinical Score

In many studies, MR parameters are tested for correlation with a clinical measure (in MS, often the Expanded Disability Status Scale, EDSS), in an attempt to investigate or demonstrate their clinical utility (or lack of it). A parameter with a high correlation coefficient is thought to be a good candidate for a surrogate MR marker of the disease in a clinical trial. In MS, low correlation coefficients r are reported (typically 0.3–0.6, sometimes 0.8; Dehmeshki *et al.*, 2001). Significance values p are also given. Correlation values are attenuated by the imperfect reliability (i.e. scatter) in the MR and clinical scores (thus even if the two measures were intrinsically perfectly correlated, the correlation plot would show scatter about the line describing this relationship). Correlation does not imply causality, only association, and the association may be weakened by the introduction of another causative factor (such as treatment). Thus a good correlation between an MR parameter and a clinical score does not necessarily imply the parameter is a good MR surrogate in a treatment trial; more evidence is needed of a direct causal relationship between the biological changes that happen in the disease and the MR parameter.

An alternative way to think of the linear regression implied in correlation is: how well can the MR parameter predict the current value of the clinical parameter? Thus a high correlation implies that the MR parameter provides a good estimate of the current clinical status (Dehmeshki *et al.*, 2001)[37] – see Figure 2.10. This establishes the relevance of the MR measure, although the ultimate goal is to predict future clinical status. The fraction of the variance in the clinical score explained by the MR parameter is r^2, and this is a useful interpretation of r.

Age and gender correlations with the principle variables can cause problems; they should be included in the correlation, as covariates (Chard *et al.*, 2002).

2.2.2.3 Clinical Scores

Clinical status can be quantified using scores. For example in MS the *EDSS* is used to measure disability. In tumours a grading system is used, based on histology of biopsy samples. In psychiatric illness, a battery of psychometric tests, including cognitive and emotional, are in use. Newly introduced MR parameters have often been judged by how well they correlate with existing clinical scores; intensive effort has gone into characterising and improving the performance of the MR parameter. In turn it has been accepted that the same intensive study should go into the clinical scores, and in fact they do have some serious shortcomings. EDSS is non-linear; mixes impairments of ambulation, fine motor skill and cognition; and has limited reproducibility (Hobart *et al.*, 2000).

[36] A study comparing MS patients and controls found a difference that was actually caused by head size (this gave a different spatial normalization and hence intensity for the two groups).

[37] In a paper by Dehmeshki (2001) a principle components analysis of MTR histograms enabled EDSS (measured values 1–8) to be predicted to within one point.

AQ: Please provide heading for Table 2.3

TABLE 2.3

AQ: Please cite Table 2.3 in text.

	Object Reported Not to Be Present (N)	Object Reported to Be Present (P)
Object absent (N)	True negative (TN)	False positive (FP)
Object present (P)	False negative (FN)	True positive (TP)

Note: The decision matrix, in receiver operating curve (ROC) formalism, is a test is to report whether an object is present or not; it could equally well be to determine whether a disease or a lesion is present or not.

AQ: Please check whether text "The decision matrix … is to report" is correct as edited; original: "Decision matrix in receiver operating curve (ROC) formalism, the test is to report"

Scores that are more appropriate are being designed. In MS the *functional composite score* (MSFC score) is increasingly popular (Fischer *et al.*, 1999; Cohen *et al.*, 2000). It consists of three components, which measure different aspects of impairment. Leg function and ambulation are measured by the timed 25-foot walk,[38] arm function by the nine-hole peg test and cognition by the Paced Auditory Serial Addition Test. Correlations of MR parameters with these individual components may be superior, since they reflect different impairments, which may occur at different times in the evolution of the disease and originate from different locations in the CNS. There is still controversy over how reliable these measurements are, since there may be learning effects in the subjects. A correlation plot may show an approximately linear dependence of a MR parameter on EDSS, but a line will often not go through the normal point (normal MR value, EDSS = 0), possibly because subclinical changes happen to the MR parameter, before any clinical disability is apparent.

AQ: Please spell out CNS at first mention

Tumour grade scoring based on sampling tissue is vulnerable to missing high-grade tissue in a heterogeneous tumour, and multicentre studies can show appreciable differences in grading between pathologists, which can limit how well MRS classifiers can work. The psychological tests are generally better, because there is a longer-standing tradition of test design in that field, although learning, floor and ceiling effects and tiredness are still important limitations.

In view of these shortcomings in clinical scores, and the complex relationship between biological changes and the ensuing clinical changes, a failure of MR parameters to closely predict these scores is hardly surprising. A more realistic test may be to look at how well the MR correlates with the biology and with scores of simple human functions.

2.2.2.4 Classification of Individual Subjects and Receiver Operating Characteristic Curves

If a measurement is performing well in separating two groups of subjects (Section 2.2.2.1), then its performance on individual subjects is worth investigating. In computer science methodology, the MR measurements may be seen as an example of a *classifier*. A classifier is a software tool for deciding which class a number of subjects belong to, based on measurements made on each subject. Classification in its simplest case only attempts to choose between two classes (binary classification). A linear discriminant is often constructed, and a threshold used to assign the class (see for example Dehmeshki *et al.*, 2002a). Such binary classification techniques have been used in spectroscopy to classify tumours into several types (Tate *et al.*, 1998). The choice of threshold in a classifier is crucial in balancing false positive and false negative errors, and receiver operating characteristic (ROC) formalism is an ideal way to view and optimise this balance. ROC analysis is also used to characterise how well a radiologist can identify a lesion on a difficult background.

Receiver operating characteristic (ROC) curves (Zweig and Campbell 1993; Altman and Bland 1994b; Armitage *et al.*, 2001; Dendy and Heaton 2002; Huo *et al.*, 2002) originate from studies to characterise screens used by radar operators, and specifically recognised that the number of objects reported would depend on how a particular operator was making their decisions. A low threshold of abnormality, i.e. reporting all objects that could possibly be real, would result in a large number of positive decisions. The proportion of actual objects detected (i.e. true positives) would be large, but at the expense of many false positives (which are in fact noise on the screen). A higher threshold (only reporting objects judged to be certainly real) would result in fewer positive decisions, more missed objects (false negatives) and fewer false positives. Thus the choice of decision threshold allows true positives to be traded off against false positives, and a particular threshold, corresponding to a particular point on the curve, can be chosen according to the relative benefit of true decisions vs the cost of false decisions. For example in a screening program, false negatives are expensive, since missing a tumour may result in death, and false positives are also costly (although less so), since they produce unnecessary worry in the subjects (Table 2.3).

In the radar screen context, an ROC curve can be generated by asking the observer to give a score with each object reported – for example 0: object absent (corresponding to a low threshold), 1: object possibly present, 2: may be present, 3: probably present, 4: almost certainly present, 5: certainly present (corresponding to a high threshold). The scores can also be defined by their anticipated probabilities (e.g. 0: <10% of being present, 1: 10%–30% of being present, etc.). The probabilities need not be correct; they just allow the observer to behave consistently. Combinations of scores are summed to give points in the ROC space corresponding to different thresholds.[39] More generally, an ROC curve can be generated from any model that predicts binary status with a certain probability, depending on one or more predictor variables, such as MR parameters. ROC curves can be fitted to analytic functions, to obtain a measure that enables ROC curves to be compared and to provide an element of smoothing (Constable *et al.*, 1995; Sorenson and Wang 1996) (Figure 2.11).

The terms *sensitivity, specificity, positive predictive value, negative predictive value, accuracy* and *prevalence* are used in the context of how well a test performs (Altman and Bland 1994a, 1994c; Dalton *et al.*, 2002). These can be defined in terms of the number of true and false positive and negative reports (see Table 2.4).

[38] Often replaced by the 10 m walk in Europe.

[39] Thus Threshold 1 corresponds to objects detected with scores 1–5 (i.e. all objects), Threshold 2 is object scores 2–5 (i.e. certainty of 'probably' or more), etc., and each threshold gives a point on the ROC curve.

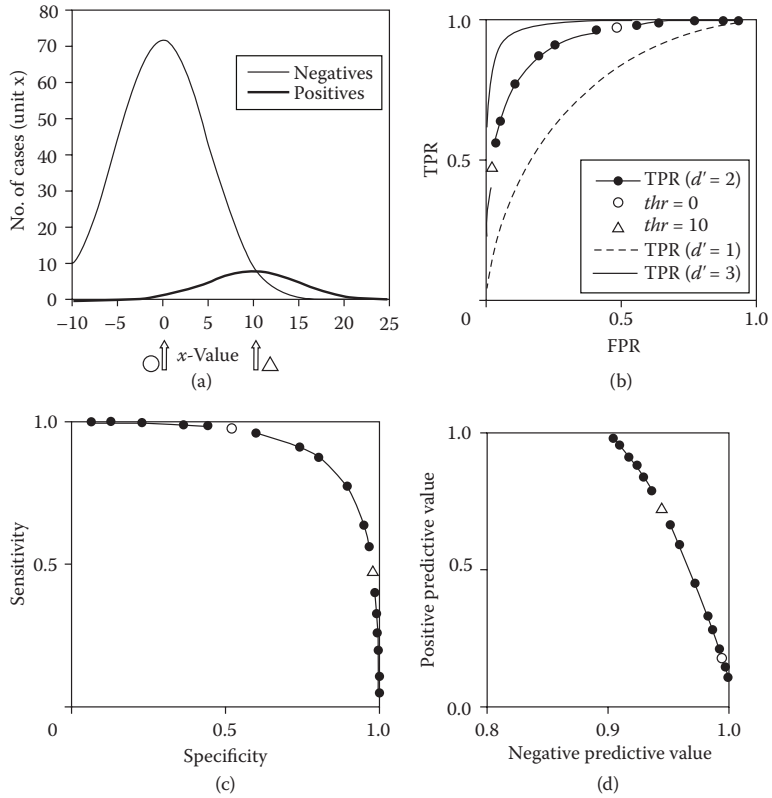


FIGURE 2.11 A receiver operating characteristic (ROC) computer simulation: (a) negative (N) and positive (P) normal distributions. There are 1000 cases, a prevalence of 1 in 10 (so 100 positive cases). Mean values are $x = 0$ and $x = 10$. Standard deviation is 5 for both groups (b) ROC curve (labelled $d' = 2$, meaning that the distributions are separated by 2 SD). Points from thresholds at the centre of each distribution are shown in parts b, c and d (*thr* = 0, centre of negatives; *thr* = 10, centre of positives). (Cases to the right-hand side of the threshold are reported positive. Thus the part of the negative tail to the right of the threshold contributes false positives, which decrease with threshold value. Conversely, the part of the positive tail to the left of the threshold gives false negatives; as the threshold is increased these increase, reducing the true positive rate.) Curves for other separations ($d' = 1$, i.e. 5 units, and $d'\_ = 3$, i.e. 15 units) are shown. The decision criterion to generate the N and P results is that an x-value greater than the threshold indicates that the sample comes from the P distribution. (c) Sensitivity vs specificity plot, which is the left–right reflection of the recursive. (d) Positive and negative predictive values both depend on the threshold (and also on the prevalence).

The false positive rate, equal to the fraction of the negative distribution that is to the right of the threshold (see Figure 2.11) is known in statistics as the significance, the α-*value*, *p-value* or the chance of a *type I error*. The false negative rate, equal to the fraction of the positive distribution that is to the left of the threshold, is the β-value, *q*-value, or the chance of a *type II error* (Haacke *et al*., 1999; Armitage *et al*., 2001). 1-β is the sensitivity or true positive rate (TPR) and is known as the power of the test (the probability of detecting a true association where one exists). Note that parameters such as the sensitivity are all estimated from samples and therefore should be reported with confidence limits.

2.2.2.5 Multiparametric Analysis

Multiparametric studies can be powerful. Several MR parameters are measured and combined in such a way to produce a better predictor of biological change or clinical outcome. If several MR parameters all correlate with clinical scores and do not correlate strongly with each other, then there is a case for measuring them all. A plea in the journal *Neurology*, on the subject of MRI techniques to monitor MS evolution (Filippi and Grossman 2002), asked that

1. Metrics from magnetization transfer MRI, diffusion-weighted MRI and proton MRS should be implemented to obtain reliable *in vivo* quantification of MS pathology.
2. Multiparametric MRI should be used in all possible clinical circumstances and trials.
3. Reproducible quantitative MR measures should ideally be used for the assessment of patients and are essential for trials.

Diagnosis at an early stage of a disease with a long and variable course (e.g. multiple sclerosis or Alzheimer's disease) and

AQ: Please provide heading for Table 2.4

TABLE 2.4

Term	Formal Definition	Formula	Comment	Dependent on Prevalence?
Sensitivity	The probability of the test finding the disease amongst those who have the disease	TP/(TP + FN)	Sensitivity characterises the false negatives.	No
Specificity	The probability of the test finding no disease amongst those who do not have the disease	TN/(TN + FP)	Specificity characterises the false positives.	No
Positive predictive value	The fraction of people with a positive test result who do have the disease	TP/(TP + FP)	Fraction of *positive* reports that are correct.	Yes
Negative predictive value	The fraction of people with a negative test result who do not have the disease	TN/(TN + FN)	Fraction of *negative* reports that are correct.	Yes
Accuracy	The fraction of test results that are correct	(TP + TN)/(TP + TN + FP + FN)	Fraction of *all reports* that are correct.	Yes
Prevalence	The fraction of the sample that has the disease	(TP + FN)/(TP + TN + FP + FN)	Distinguish from incidence.[a]	–
True positive rate	Fraction of positive cases that are detected	TP/(TP + FN)	Equals sensitivity. The fractional area under the positive curve that is to the right of the threshold.	No
False positive rate	Fraction of negative cases that are reported positive	FP/(TN + FP)	Equals 1-specificity. The fractional area under the negative curve that is to the right of the threshold.	No

Note: Radiological terms summarising the performance of a test; in this case the test is to find out whether subjects have a disease or not. Negative indicates that they do not, positive that they do. Sensitivity and specificity define the performance of the test (regardless of prevalence), whilst predictive value and accuracy depend on prevalence (i.e. what fraction of the sample has the disease). Thus the latter quantities are very different for an asymptomatic screened population and a symptomatic hospital population.

[a] Thus the prevalence is the fraction of people that have a particular disease at any one time. The *incidence* is the number of *new cases* of disease that arise in a given period (usually 1 year) for a given size of population.

prediction of disease progression can also be viewed as a multiparametric problem. Early symptoms, MRI and other biochemical data are available; their combination can be optimised to provide the best prediction of future clinical status, measured according to various clinical scores. The optimisation can be driven by criteria such as sensitivity, specificity and predictive value (see Table 2.4).

The effects of disease are similarly multidimensional and need several scores to properly characterise them, covering both clinical symptoms and measurable biological changes. To assemble multiparametric MR data, features can be extracted from histograms of each MR parameter and from lesion values. Statistical techniques such as multiple linear discriminant analysis and cluster analysis are appropriate (Tintore *et al.*, 2001). Multiparametric analysis is discussed further in Chapter 18.

References

Alecci M, Collins CM, Smith MB, Jezzard P. Radio frequency magnetic field mapping of a 3 Tesla birdcage coil: experimental and theoretical dependence on sample properties. Magn Reson Med 2001; 46(2): 379–85.

Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ 1994a; 309(6947): 102.

Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994b; 309(6948): 188.

Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ 1994c; 308(6943): 1552.

Andersen AH. On the Rician distribution of noisy MRI data. Magn Reson Med 1996; 36(2): 331–3.

Armitage P, Matthews JNS, Berry G. *Statistical Methods in Medical Research*. Blackwell; 2001.

Barker GJ, Simmons A, Arridge SR, Tofts PS. A simple method for investigating the effects of non-uniformity of radiofrequency transmission and radiofrequency reception in MRI. Br J Radiol 1998; 71(841): 59–67.

Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310(6973): 170.

Bottomley PA, Andrew ER. RF magnetic field penetration, phase shift and power dissipation in biological tissue: implications for NMR imaging. Phys Med Biol 1978; 23(4): 630–43.

Boudreau M, Tardif CL, Stikov N, Sled JG, Lee W, Pike GB. B1 mapping for bias-correction in quantitative T1 imaging of the brain at 3T using standard pulse sequences. J Magn Reson Imaging 2017.

AQ: Boudrea et al. (2017): Please provide volume number and page range.

Brown RW, Cheng N, Haacke EM, Thompson MR, Venkatesan R. *Magnetic resonance imaging: Physical principles and sequence design* (2nd edition): Wiley-Blackwell; 2014.

AQ: Brown et al. (2014): Please provide publisher location.

Cercignani M, Inglese M, Siger-Zajdel M, Filippi M. Segmenting brain white matter, gray matter and cerebro-spinal fluid using diffusion tensor-MRI derived indices. Magn Reson Imaging 2001; 19(9): 1167–72.

Chard DT, Griffin CM, Parker GJ, Kapoor R, Thompson AJ, Miller DH. Brain atrophy in clinically early relapsing-remitting multiple sclerosis. Brain 2002; 125(Pt 2): 327–37.

Cohen JA, Fischer JS, Bolibrush DM, Jak AJ, Kniker JE, Mertz LA, et al. Intrarater and interrater reliability of the MS functional composite outcome measure. Neurology 2000; 54(4): 802–6.

Collins CM, Liu W, Swift BJ, Smith MB. Combination of optimized transmit arrays and some receive array reconstruction methods can yield homogeneous images at very high frequencies. Magn Reson Med 2005; 54(6): 1327–32.

Collins CM, Smith MB. Signal-to-noise ratio and absorbed power as functions of main magnetic field strength, and definition of "90 degrees " RF pulse for the head in the birdcage coil. Magn Reson Med 2001; 45(4): 684–91.

Collins CM, Wang Z. Calculation of radiofrequency electromagnetic fields and their effects in MRI of human subjects. Magn Reson Med 2011; 65(5): 1470–82.

Collins CM, Yang QX, Wang JH, Zhang X, Liu H, Michaeli S, et al. Different excitation and reception distributions with a single-loop transmit-receive surface coil near a head-sized spherical phantom at 300 MHz. Magn Reson Med 2002; 47(5): 1026–8.

Constable RT, Skudlarski P, Gore JC. An ROC approach for evaluating functional brain MR imaging and postprocessing protocols. Magn Reson Med 1995; 34(1): 57–64.

Cunningham CH, Pauly JM, Nayak KS. Saturated double-angle method for rapid B1+ mapping. Magn Reson Med 2006; 55(6): 1326–33.

Dalton CM, Brex PA, Miszkiel KA, Hickman SJ, MacManus DG, Plant GT, et al. Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis. AnnNeurol 2002; 52(1): 47–53.

Dehmeshki J, Barker GJ, Tofts PS. Classification of disease subgroup and correlation with disease severity using magnetic resonance imaging whole-brain histograms: application to magnetization transfer ratios and multiple sclerosis. IEEE Trans Med Imaging 2002a; 21(4): 320–31.

Dehmeshki J, Ruto AC, Arridge S, Silver NC, Miller DH, Tofts PS. Analysis of MTR histograms in multiple sclerosis using principal components and multiple discriminant analysis. Magn Reson Med 2001; 46(3): 600–9.

Dehmeshki J, Van Buchem MA, Bosma GP, Huizinga TW, Tofts PS. Systemic lupus erythematosus: diagnostic application of magnetization transfer ratio histograms in patients with neuropsychiatric symptoms--initial results. Radiology 2002b; 222(3): 722–8.

AQ: Dendy and Heaton (2002): Please provide publisher location.

Dendy PP, Heaton B. *Physics for diagnostic radiology* Institute of Physics; 2002.

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(3): 622–30.

Duan Q, van Gelderen P, Duyn J. Improved Bloch-Siegert based B1 mapping by reducing off-resonance shift. NMR Biomed 2013; 26(9): 1070–8.

Edelstein WA, Bottomley PA, Pfeifer LM. A signal-to-noise calibration procedure for NMR imaging systems. Med Phys 1984; 11(2): 180–5.

Fernandez-Seara MA, Song HK, Wehrli FW. Trabecular bone volume fraction mapping by low-resolution MRI. Magn Reson Med 2001; 46(1): 103–13.

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.

Goerner FL, Clarke GD. Measuring signal-to-noise ratio in partially parallel imaging MRI. Med Phys 2011; 38(9): 5049–57.

Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magn Reson Med 1995; 34(6): 910–4.

Haacke EM, Brown RW, Thompson MR, Venkatesan R. *Magnetic Resonance Imaging*; 1999Physical principles and sequence design.

Haynes BI, Dowell NG, Tofts PS. Measuring scan-rescan reliability in quantitative brain imaging reveals instability in an apparently healthy imager and improves statistical power in a clinical study. ISMRM annual scientific meeting; 2010; Stockholm; 2010. p. 2999.

Henkelman RM. Measurement of signal intensities in the presence of noise in MR images. Med Phys 1985; 12(2): 232–3.

Henkelman RM. Erratum: measurement of signal intensities in the presence of noise [Med. Phys. 12, 232 (1985)]. Med Phys 1986; 13(4): 544.

Hetherington HP, Chu WJ, Gonen O, Pan JW. Robust fully automated shimming of the human brain for high-field 1H spectroscopic imaging. Magn Reson Med 2006; 56(1): 26–33.

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

Hoult DI. The NMR receiver: a description and analysis of design. Progress in NMR Spectroscopy 1978; 12: 41–77.

Hoult DI. The principle of reciprocity in signal strength calculations – a mathematical guide. Concepts in Magnetic Resonance 2000; 12: 173–87.

Hoult DI, Richards RE. The signal-to-noise ratio of the nuclear magnetic resonance experiment. J Magn Reson 1976; 24: 71–85.

Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. Radiology 2002; 224(2): 560–8.

Hurley SA, Yarnykh VL, Johnson KM, Field AS, Alexander AL, Samsonov AA. Simultaneous variable flip angle-actual flip angle imaging method for improved accuracy and precision of three-dimensional T1 and B1 measurements. Magn Reson Med 2012; 68(1): 54–64.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R. Image distortion correction in fMRI: a quantitative evaluation. NeuroImage 2002; 16(1): 217–40.

Ibrahim TS. Analytical approach to the MR signal. Magn Reson Med 2005; 54(3): 677–82.

Ibrahim TS, Lee R, Baertlein BA, Abduljalil AM, Zhu H, Robitaille PM. Effect of RF coil excitation on field inhomogeneity at ultra high fields: a field optimized TEM resonator. Magn Reson Imaging 2001; 19(10): 1339–47.

Insko EK, Bolinger L. Mapping the radiofrequency field. J Magn Reson series A 1993; 103: 82–5.

AQ: Jezzrad (2002): Please provide publisher location and page range.

Jezzard P. Physical basis of spatial distortions in Magnetic Resonance Images. In: Isaac B, editor. *Handbook of Medical Imaging*: Academic Press; 2002.

Jin J, Liu F, Zuo Z, Xue R, Li M, Li Y, et al. Inverse field-based approach for simultaneous B_1 mapping at high fields - a phantom based study. J Magn Reson 2012; 217: 27–35.

Karlsen OT, Verhagen R, Bovee WM. Parameter estimation from Rician-distributed data sets using a maximum likelihood estimator: application to T1 and perfusion measurements. Magn Reson Med 1999; 41(3): 614–23.

Lemieux L, Barker GJ. Measurement of small inter-scan fluctuations in voxel dimensions in magnetic resonance images using registration. Med Phys 1998; 25(6): 1049–54.

Lutti A, Hutton C, Finsterbusch J, Helms G, Weiskopf N. Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. Magn Reson Med 2010; 64(1): 229–38.

Michaelis T, Merboldt KD, Bruhn H, Hanicke W, Frahm J. Absolute concentrations of metabolites in the adult human brain in vivo: quantification of localized proton MR spectra. Radiology 1993; 187(1): 219–27.

Miller AJ, Joseph PM. The use of power images to perform quantitative analysis on low SNR MR images. Magn Reson Imaging 1993; 11(7): 1051–6.

Moerland MA, Beersma R, Bhagwandien R, Wijrdeman HK, Bakker CJ. Analysis and correction of geometric distortions in 1.5 T magnetic resonance images for use in radiotherapy treatment planning. Phys Med Biol 1995; 40(10): 1651–4.

Morrell GR. A phase-sensitive method of flip angle mapping. Magn Reson Med 2008; 60(4): 889–94.

Morrell GR, Schabel MC. An analysis of the accuracy of magnetic resonance flip angle measurement methods. Phys Med Biol 2010; 55(20): 6157–74.

Murphy BW, Carson PL, Ellis JH, Zhang YT, Hyde RJ, Chenevert TL. Signal-to-noise measures for magnetic resonance imagers. Magn Reson Imaging 1993; 11(3): 425–8.

Nehrke K, Börnert P. DREAM – a novel approach for robust, ultrafast, multislice B_1 mapping. Magn Reson Med 2012; 68(5): 1517–26.

Park DJ, Bangerter NK, Javed A, Kaggie J, Khalighi MM, Morrell GR. A statistical analysis of the Bloch-Siegert B1 mapping technique. Phys Med Biol 2013; 58(16): 5673–91.

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(5): 838–45.

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(1): 27–41.

Pohmann R, Scheffler K. A theoretical and experimental comparison of different techniques for B_1 mapping at very high fields. NMR Biomed 2013; 26(3): 265–75.

Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magn ResonMed 1993; 30(6): 672–9.

Provencher SW. Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR Biomed 2001; 14(4): 260–4.

Sabati M, Maudsley AA. Fast and high-resolution quantitative mapping of tissue water content with full brain coverage for clinically-driven studies. Magn Reson Imaging 2013; 31(10): 1752–9.

Sacolick LI, Sun L, Vogel MW, Dixon WT, Hancu I. Fast radiofrequency flip angle calibration by Bloch-Siegert shift. Magn Reson Med 2011; 66(5): 1333–8.

Sacolick LI, Wiesinger F, Hancu I, Vogel MW. B1 mapping by Bloch-Siegert shift. Magn Reson Med 2010; 63(5): 1315–22.

Sbrizzi A, Raaijmakers AJ, Hoogduin H, Lagendijk JJ, Luijten PR, van den Berg CA. Transmit and receive RF fields determination from a single low-tip-angle gradient-echo scan by scaling of SVD data. Magn Reson Med 2014; 72(1): 248–59.

Sijbers J, den Dekker AJ, Van Audekerke J, Verhoye M, Van Dyck D. Estimation of the noise in magnitude MR images. Magn Reson Imaging 1998; 16(1): 87–90.

Sled JG, Pike GB. Standing-wave and RF penetration artifacts caused by elliptic geometry: an electrodynamic analysis of MRI. IEEE Trans Med Imaging 1998; 17(4): 653–62.

Sled JG, Pike GB. Correction for B(1) and B(0) variations in quantitative T(2) measurements using MRI. Magn Reson Med 2000; 43(4): 589–93.

Sorenson JA, Wang X. ROC methods for evaluation of fMRI techniques. Magn Reson Med 1996; 36(5): 737-44.

Stollberger R, Wach P. Imaging of the active B1 field in vivo. Magn Reson Med 1996; 35(2): 246–51.

Tate AR, Griffiths JR, Martinez-Perez I, Moreno A, Barba I, Cabanas ME, et al. Towards a method for automated classification of 1H MRS spectra from brain tumours. NMR Biomed 1998; 11(4–5): 177–91.

Tintore M, Rovira A, Brieva L, Grive E, Jardi R, Borras C, et al. Isolated demyelinating syndromes: comparison of CSF oligoclonal bands and different MR imaging criteria to predict conversion to CDMS. MultScler 2001; 7(6): 359–63.

Tofts PS. Optimal detection of blood-brain barrier defects with Gd-DTPA MRI-the influences of delayed imaging and optimised repetition time. Magn Reson Imaging 1996; 14(4): 373–80.

Tofts PS, Kermode AG, MacManus DG, Robinson WH. Nasal orientation device to control head movement during CT and MR studies. J Comput Assist Tomogr 1990; 14(1): 163–4.

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, Wray S. Noninvasive measurement of molar concentrations of 31P metabolites in vivo, using surface coil NMR spectroscopy. Magn Reson Med 1988; 6(1): 84–6.

Tozer DJ, Tofts PS. Removing spikes caused by quantization noise from high-resolution histograms. Magn Reson Med 2003; 50(3): 649–53.

Tubridy N, McKinstry CS. Neuroradiological history: Sir Joseph Larmor and the basis of MRI physics. Neuroradiology 2000; 42(11): 852–5.

Venkatesan R, Lin W, Haacke EM. Accurate determination of spin-density and T1 in the presence of RF-field inhomogeneities and flip-angle miscalibration. Magn Reson Med 1998; 40(4): 592–602.

Volz S, Nöth U, Deichmann R. Correction of systematic errors in quantitative proton density mapping. Magn Reson Med 2012; 68(1): 74–85.

Volz S, Nöth U, Rotarska-Jagiela A, Deichmann R. A fast B1-mapping method for the correction and normalization of magnetization transfer ratio maps at 3 T. NeuroImage 2010; 49(4): 3015–26.

Watanabe H, Takaya N, Mitsumori F. Non-uniformity correction of human brain imaging at high field by RF field mapping of B1+ and B1–. J Magn Reson 2011; 212(2): 426–30.

Weisskoff RM. Simple measurement of scanner stability for functional NMR imaging of activation in the brain. Magn Reson Med 1996; 36(4): 643–5.

Wheeler-Kingshott CA, Parker GJ, Symms MR, Hickman SJ, Tofts PS, Miller DH, et al. ADC mapping of the human optic nerve: increased resolution, coverage, and reliability with CSF-suppressed ZOOM-EPI. Magn Reson Med 2002; 47(1): 24–31.

Whisenant JG, Dortch RD, Grissom W, Kang H, Arlinghaus LR, Yankeelov TE. Bloch-Siegert B1-Mapping improves accuracy and precision of longitudinal relaxation measurements in the breast at 3 T. Tomography 2016; 2(4): 250–9.

Yarnykh VL. Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magn Reson Med 2007; 57(1): 192–200.

Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. ClinChem 1993; 39(4): 561–77.