

Introduction to quantitative analysis

ISMRM 2006 Seattle

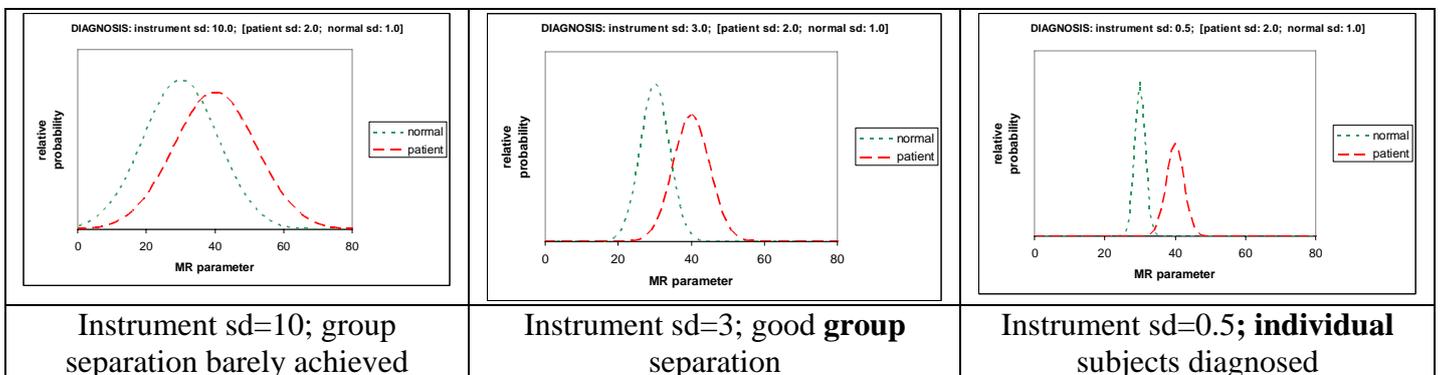
[Paul Tofts](#) PhD

Version December 8th 2005

Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK

1. Overview and concepts

- a. On *completion*, you should be able to:
 - i. Describe 5 basic principles of good methodology in qA
 - ii. Describe 5 sources of variation in qA and methods to reduce them
- b. Look forward to:
 - i. talk 2 - *Mapping* – Gareth Barker
 - ii. talk 3 – statistical analysis – John Petkau
- c. Image analysis and quantification. Image analysis is near the end of the *measurement process*, and without attention to the whole, results may be almost meaningless. Hence concepts of quantification, qMR, are needed.
- d. *qMR* uses the paradigm of a scientific instrument. Long history. Well worked out concepts (accuracy, precision, uncertainty etc). The ‘happy-snappy MRI camera’ evolves to measure clinically relevant biological quantities. Expect the same performance as when measuring e.g. body height or weight.
- e. *Applications* are in diagnosis, monitoring disease progression, and measuring response to treatment (trials, and customised treatment).
- f. *Images* are collected with various contrasts (weightings) e.g. T₁ or D. From these, *maps* are calculated showing values of a MR parameter (e.g. T₁ or ADC).
- g. *Partial volume* effects (voxel too large or elongated) often prevent us measuring a pure tissue (e.g. grey matter). Use small isotropic voxels where possible (compromise with SNR).
- h. *Study design*:
 - i. Beware ‘stamp collecting’; clarify objective and hypothesis from the start. Read the literature!
 - ii. Use interleaved controls in a serial study, to control for scanner drift.
 - iii. Talk to a statistician



Example of how good instrument performance can affect clinical possibilities

The measurement process - data collection

- i. Awareness of *data collection* (+ sources of variation) guides study design, image analysis and interpretation, and may lead to re-design of data collection
- j. Subject *positioning*
 - i. *Inter-scan* movement often corrected using spatial registration (more later)
 - ii. *Intra-scan movement* causes
 1. artefacts (usually in phase encode direction)
 2. loss of signal (to air – wind grey window down and look!).
- k. Scanner *Auto Pre Scan* procedure
 - i. Set transmitter output
 - ii. Set receiver gain
 - iii. Set centre frequency
 - iv. Shim
 - v. Can be manually over-ridden for serial scans (e.g. dynamic Gd enhancement)
- l. *NMR signal* is proportional
 - i. to proton density (spins per ml) AND
 - ii. $1/(\text{absolute temperature})$. From refrigerator (278K) to brain (310K), signal decreases 10% (!)
- m. *static field* B_0 – local susceptibility gradients near air-tissue interfaces cause shift and dropout, especially in gradient echo and EPI.
- n. *Static field gradients*
 - i. larger values can overcome local susceptibility gradients.
 - ii. Eddy currents can distort diffusion imaging and spectroscopy
- o. *Radiofrequency transmit field* B_1
 - i. B_1 and flip angle (FA) errors are the largest source of error in qMR. Crucial in e.g. Magnetisation Transfer, T_1 measurement.
 - ii. *Absolute value* is set by APS.
 - iii. *RF nonuniformity* causes B_1 to vary with position
 - iv. $B_1(r)$ can be mapped; worst for small transmit/receive coil (e.g. head coil); best for body transmit coil.
 - v. *Residual nonuniformity* caused by subject; increases with field (3T..).
- p. *Slice selection*
 - i. In 2D sequences, FA varies across slice profile, and slice profile varies with position (if B_1 is nonuniform). Can degrade quantification. Can be corrected using knowledge of selective pulse and $B_1(r)$
 - ii. $B_1(r)$ mapping is a growing subject (later talk by Geoff Parker).
 - iii. 3D sequences preferred (harder broadly-selective RF pulse).
- q. *Receive sensitivity*
 - i. *Coil loading* (depends on subject). Electrical losses reduce Q.
 - ii. Coil loading can in principle (usually) be found from the transmitter setting using the *principle of reciprocity*:
(Receive sensitivity) x (transmitter output for a 90° pulse) = constant (independent of loading)
 - iii. *Receiver gain* (adjusted by APS).
 - iv. *Image noise* – measure in surrounding air ($sd_{\text{air}} = 0.6 sd_{\text{subject}}$). SNR can be increased (at expense of imaging time, spatial resolution, unwanted T_1 -weighting etc)
 - v. *Receive nonuniformity*
 1. can degrade qualitative appearance (*close fitting coils* for high SNR)
 2. often does not degrade qMR, since *ratios* of images are used

3. exceptions: PD and spectroscopy (*absolute* signal is important).
 4. Sometimes corrected in software: usually only qualitative
 5. Can be corrected using phantom scan or B_1 map (but...)
- r. *Image registration* – often done
- i. Serial studies – looking for subtle changes over time
 - ii. Subtraction (e.g. Gd) – subtle changes
 - iii. Group mapping – subtle differences between individual and a group
 - iv. See talk 3 - Mark Jenkinson

2. The measurement process - image analysis

- a. Types of image analysis
- i. Region of Interest (ROI)
 1. *simplest*: locate in various parts of the body.
 2. *measure* mean value, sd and number of pixels (area in mm^2)
 3. Can be 2D (single slice) or 3D (VOI, several slices).
 4. manual or semi-automatic *generation*.
 5. *Check adjacent slices* for partial volume errors (if measuring mean)
 6. *Location of ROI* is source of variation in repeats (especially near edges, and if defined manually). Observer should be blinded to subject status.
 7. Many studies on *intra-observer and inter-observer* variation (within-observer vs between observer). E.g. Multiple Sclerosis(MS) (visible) lesion load.
 - a. Re-analyse same image dataset. Analyse discrepancies, define rules.
 - b. Scan-rescan: more realistic. Includes variation from whole measurement process.
 8. use an *independent 'structural image'* to define ROI's. If these are defined on parametric map (e.g. MTR), then mean values of parameter can be biased.
 - ii. Histogram analysis
 1. All voxels in a volume (e.g. whole brain or tumour) are counted, and a frequency distribution plotted. MORE in NEXT SECTION.
 2. Spatial discrimination lost; however very sensitive to subtle diffuse abnormality (e.g. Normal Appearing White Matter in MS).
 - iii. Voxel-based group mapping (VBGM)
 1. ROI problem: where to place the ROI? Which location to test? Multiple comparisons, Bonferroni.
 2. Histogram problem: spatial information lost. Abnormality may be localised (but where?)
 3. VBGM: Every location (size about 1cm cubed) is tested for abnormality, without prior choices.
 4. Images from a group of subjects are registered to a standard template (more later?)
 5. Conceptually between ROI and histogram
- b. Types of statistical analysis
- i. Complex; often done badly. Talk to statisticians early on.
 - ii. Often the end point of a long study; get it right!
 - iii. Control for age and gender during recruitment and analysis (covariate)
 - iv. Inspect the data (scatter plots)

- v. *Group comparisons*
 - 1. *t-test* (parametric data).
 - 2. Beware *multiple comparisons* (Bonferroni correction? Too harsh?)
 - 3. *Fishing expedition* vs. hypothesis testing
 - 4. give *confidence limits* as well as p-values.
 - a. ‘no difference was found, although a previous study had shown one’ - WHY?
 - b. ‘the 95%CI for the group difference was -0.5 to +2.5’
 - c. ‘in a previous study the 95%CI was 0.5 to 1.5 (better precision) OR
 - d. ‘in a previous study the 95%CI was 4.5 to 7.5 (larger separation)
- vi. *negative results* (in group comparison)
 - 1. there may be only a *small between-group* effect OR
 - 2. the *within group-effect* may be large (reduce it? e.g. age dependency) OR
 - 3. the *instrumental variation* may be (unnecessarily) LARGE
 - 4. thus *better technique* could change a negative result to positive.
 - 5. Analyse and quantify *sources of variation*
- vii. *false positive result*
 - 1. e.g. patients before a *scanner upgrade*, controls after (!)
- viii. *correlations*
 - 1. Widely used, but...
 - 2. Correlating to a ‘gold standard’ e.g. clinical score, pathological score, soon uncovers corrosion. Between-rater disagreement. Relation of biology to clinical status?
 - 3. Failure of a MR parameter to correlate with a clinical score does not imply it’s a ‘bad’ parameter. Correlation with biology (biopsy or post mortem samples) is ?more meaningful.
- ix. *Classification*
 - 1. Can individual subjects be well classified by a MR parameter?
 - 2. More useful than group comparisons, if the parameter is specific and sensitive enough. Diagnosis, monitor treatment etc.
 - 3. example from NPSLE and MTR histograms (Dehmeshki).
 - 4. trade off sensitivity (no false negatives) vs. specificity (no false positives). Visualised with Receiver Operating Characteristic (ROC) curves.

3. Histograms

- a. *MR parameter* measured over large volume (e.g. brain or white matter)
- b. *Frequency distribution*
- c. very *subtle diffuse changes* can be detected – e.g. NAWM in MS
- d. process includes:
 - i. MR parameter measurement – MR physics - elsewhere
 - ii. Image segmentation – talk 4 – Tim Cootes
 - iii. Histogram generation – examined in detail
- e. Step 1: *despiking*
 - i. raw image datasets are usually stored as integers, with sufficient resolution so that rounding errors do not contribute to image noise.
 - ii. These are converted to floating point numbers.
 - iii. Parameter (e.g. MTR or T_1) is calculated

- iv. Histogram spikes can be produced, arising from the discrete distribution in the raw image data.
- v. Random noise (-0.5 to +0.5) added to raw images avoids this.
- f. Step 2: histogram *architecture*
 - i. Define bin width
 - 1. too large and peak structure is lost; features of peak height and location affected.
 - 2. too small and too much noise (though best to use small bin and smooth)
 - ii. define bin borders
 - 1. e.g. bin at 30 could be 29.51 to 30.49 (i.e. central labelling) or 30.00 to 30.99 (edge labelling)
 - 2. beware parameter value falling on bin interface
 - 3. small bins reduce this problem
- g. step 3: *smoothing*
 - i. e.g. median smoothing of small bins
- h. step 4: *normalisation*
 - i. 'partial normalisation' divide by total no. of voxels, to account for e.g. varying brain sizes
 - ii. 'full normalisation' divide by bin width (and multiply by 100%)
 - 1. gives peak height independent of bin width

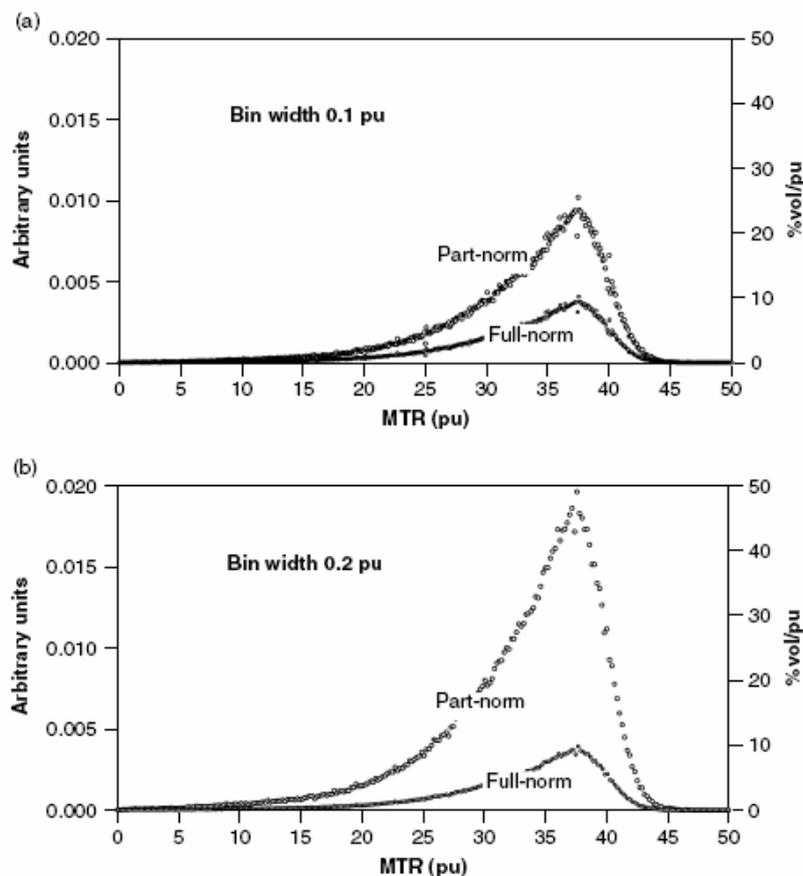


Figure 18.7. Partly normalized and fully normalized MTR histograms. The partly normalized histograms were calculated by dividing each absolute histogram value by the sum of the values (without taking account of bin width); they have a peak height that is dependent on the bin width (in fact inversely proportional to it). The fully normalized histograms were calculated taking account of bin width [Equation (18.4)]; they are independent of bin width, with a peak height of about 10 % vol/pu, a peak width (full-width half-maximum) of 10 pu, and an area of 100 % vol

2. units are then %vol per [MR parameter unit] e.g. % vol/ms for T₁
 3. total area under histogram curve is 100%
- i. step 5: *feature extraction*
 - i. peak height (decreases if histogram broadens)
 - ii. peak location
 - iii. alternatively, fit function to peak(s) to determine height and location
 - iv. percentiles (i.e. area in tails), e.g. 25th percentile
 - v. mean value in segmented volume (no need for a histogram!)
 - vi. Principle Component Analysis and Linear Discriminant Analysis
 - j. EXAMPLE of *partial and full-normalisation*

4. Quality assurance

- a. QA of MR measurements can establish
 - i. short-term stability
 - ii. accuracy (closeness to truth)
 - iii. long term stability (for long serial studies)
- b. QA can use normal controls
 - i. + stable
 - ii. + realistic
 - iii. - truth not known
 - iv. - need more resources
- c. QA can use phantoms (test objects)
 - i. + less resources
 - ii. + truth is known
 - iii. – poor stability – temperature 2-3%/°C
 - iv. – poor stability – chemical decay
 - v. – often not realistic
- d. Sources of random variation
 - i. Biological – normal variation (e.g. blood flow)
 - ii. Data collection (see ‘the measurement process’)
 1. subject positioning (head in coil, slices in head)
 2. which radiographer? (technologist)
 3. pre-scan procedure
 4. Gd injection procedure
 5. subject movement (macroscopic)
 6. subject movement (cardiac, CSF, breathing)
 7. image noise
 - iii. Image analysis
 1. ROI placement
 2. segmentation
 3. within-observer and between-observer effects
 - a. training (set of rules)
 - b. observation conditions
 - c. psychological state (the maternity syndrome)
 - d. (none with automatic system)
- e. Why does random instrumental error matter?
 - i. If it is comparable with the biological effects being measured, it can make a study result negative
 - ii. In a cross-sectional study, it should be << natural within-group variance
 1. ‘No difference was found, contrary to the findings by other groups’

- iii. In a serial study it should be \ll biological within-subject change over time
 - 1. 'there was no significant change over a period of one year'
- f. Measurement of random error
 - i. Bland-Altman:
 - 1. pairs of repeated measurements
 - 2. give sd of a single measurement
 - 3. and 95% confidence limits in a single measurement
 - ii. Intra-class Correlation Coefficient (ICC) (=reliability)
 - 1. = (variance between subjects) / (var between + var within)
 - 2. typically 0.9
 - 3. shows fraction of measured variance that comes from within-group differences (as opposed to instrument)
 - 4. sets instrumental random error in context BUT
 - 5. depends on subject group (e.g. good for patients, bad for controls)
- g. Sources of systematic error
 - i. Biological – hope controls are stable
 - ii. Data collection
 - 1. B_1 error (including nonuniformity)
 - 2. slice profile, partial volume
 - 3. operator training
 - 4. hardware/software upgrade
 - iii. Image analysis
 - 1. observer training
 - 2. software upgrades
- h. does systematic error matter?
 - i. In short-term single-centre study – NO
 - ii. In multi-centre studies – YES
 - iii. In long studies – PROBABLY
 - 1. systematic error (if present, and not understood) can vary with centre and with time.
 - iv. Example of how medium term changes in systematic error wrecked a long-term study (fictitious)
 - 1. figure (from qMR book¹)

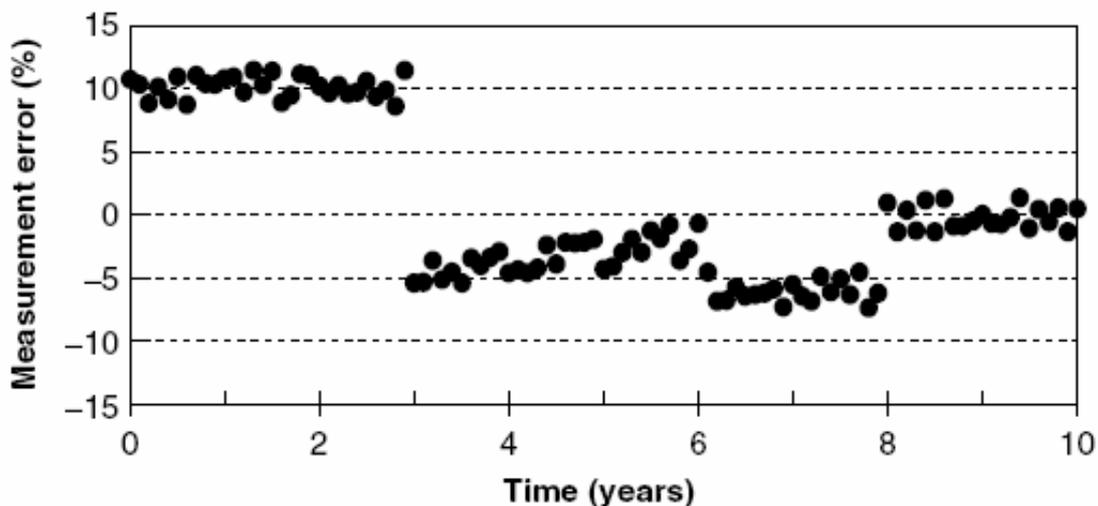


Figure 3.5. Long-term precision is dominated by instability in the systematic error. Simulation of fictional change in measurement error over time, during a longitudinal study. Short-term precision is good, and a study completed in the first 3 years is unaffected by the large systematic error (i.e. poor accuracy). A major upgrade at year 3 dramatically changes the systematic error. A subtle drift in values takes place, followed by two more step changes, at the times of operator change and a minor upgrade. At year 8 the sources of systematic error are finally identified and removed, giving a system that should provide good accuracy and hence long-term precision for many years

2. A realistic phantom for MS lesions (OCCA) showed large errors in estimated volume, dependent on the observer and the lesion contrast.
3. Understanding inaccuracy is good MR physics!
 - i. Understanding random error
 - i. It's often assumed that image noise determines this; however not so (?).
 - ii. By identifying and measuring sources of random error, we can refine the measurement process, optimise parameters (e.g. TR, NEX), minimise random error for a given examination time.
 - j. Phantoms exist for many MR quantities (but not all)
 - i. Volume – easy – geometric objects, Perspex and water
 - ii. T_1 T_2 – doped agarose gels. Ni T_1 is temperature independent
 - iii. ADC – alkanes (although T_1 T_2 too long)
 - iv. DTI – hard. Need bundles of small fibres.
 - v. MT – BSA
 - vi. Temperature control and monitoring possible to 0.2°C

5. Summary

- a. By being aware of all stages in the measurement process, and where sources of variation and error can be present, the quality of the results from quantitative image analysis is improved. As a result studies are more likely to give meaningful and useful results.
- b. Some principles:
 - i. Understand systematic error in data collection, especially RF nonuniformity
 - ii. Place ROI's carefully
 - iii. Understand statistics
 - iv. Use QA in phantoms and controls
 - v. Understand and minimise sources of random error in data collection and image analysis

7. Additional material:

1. The material in this presentation is based on chapters ¹⁻⁴ in the book *Quantitative MRI of the Brain*.
2. The website qmri.org contains additional material on qMR, and will include an up-to-date version of this presentation.
3. Other helpful books are listed in², and include the comprehensive book by Haacke et al⁵

Reference List

1. Tofts PS. QA: Quality assurance, accuracy, precision and phantoms (chapter 3). In: Paul Tofts, editor. *Quantitative MRI of the brain: measuring change caused by disease*. Chichester: John Wiley, 2003: 55-81.
2. Tofts PS. Concepts: Measurements and MR (chapter 1). In: Paul Tofts, editor. *Quantitative MRI of the brain: measuring change caused by disease*. Chichester: John Wiley, 2003: 3-15.

3. Tofts PS. The measurement process: MR data collection and image analysis (chapter 2). In: Paul Tofts, editor. Quantitative MRI of the brain: measuring change caused by disease. Chichester: John Wiley, 2003: 17-54.
4. Tofts PS, Davies GR, Dehmshki J. Histograms: measuring subtle diffuse disease (chapter 18). In: Paul Tofts, editor. Quantitative MRI of the brain: measuring change caused by disease. Chichester: John Wiley, 2003: 581-610.
5. Haacke EM, Brown RW, Thompson MR, Venkatesan R. Magnetic Resonance Imaging. physical principles and sequence design. New York: Wiley-Liss, 1999.