# Measuring scan-rescan reliability in quantitative brain imaging reveals instability in an apparently healthy imager and improves statistical power in a clinical study.

**B. I. Haynes[1], N. G. Dowell[1], and P. S. Tofts[1]**
[1]Clinical Imaging Sciences Centre, Brighton and Sussex Medical School, Brighton, United Kingdom

**Hypotheses**: 1: measuring and optimising scan-rescan repeatability can increase the statistical power of a clinical study. 2: published values of normal variance will include a variable contribution from instrumental variation

**Background**: Repeatability of MTR and ADC brain histograms of healthy volunteers in our centre showed disturbingly large differences, even though the scanner was producing high quality images. MTR mapping relies on short-term stability, since the map is formed from the difference between two images. Such instrumental variation could mask small between-group differences in a cross-sectional study or intra-individual differences in a serial study and could also explain inconsistent findings in published studies. There are almost no published data on short- or long-term within-subject reproducibility, although the factors which contribute to multi-centre variance in MTR and ADC are understood and probably similar [1,2].

**Methods:** *1. Modelling dependence of statistical power on instrumental variance*. A simulation for a cross-sectional study was built, based on realistic parameters for MTR in normal-appearing white matter (Control group: mean=37 pu, intrinsic sd=0.4. Patient group: mean=36.5 pu, sd=0.65)[1,4-10]. The total variance in each group was calculated by adding intrinsic (biological) and instrumental variances – ranging from 0-1 pu. A power calculation was performed to gauge the sample size needed to observe the group difference at each instrumental sd using the following values: α=0.05, two tailed, power=0.8, group sizes equal.
*2. Measurement of normal range*. Nine healthy controls (age 36±13 (mean±sd)) were imaged; after the transmitter boards were changed a further nine were imaged (age 38±10). The 95% confidence limits (95%CL) in the estimate of sd was calculated from the Chi-squared distribution [3].
*3. Measurement of instrumental variation in controls*. Pairs of repeated measurements were made in healthy volunteers (n=13). The repeats were carried out during the same scanning session, to evaluate instant (i.e. short-term) repeatability. Bland-Altmann analysis was used to estimate the sd of a single measurement (ISD = RMS value of differences/1.4).
*4. Measurement of instrumental variation using a phantom*. A cylindrical water phantom (doped with $NiSO_4$) was imaged repeatedly after suspicions were aroused from volunteer scanning that machine repeatability was poor. A 3D FLASH sequence was used (acquisition time 6 mins) repeated for 20 hours. Mean signal was calculated from a ROI placed in the central slice of each acquisition. This process was automated using a MATLAB program. Running difference was calculated as percentage difference in signal at adjacent time points.
*5. Published values of normal group variation*. The measurement of white matter MTR is established as a useful biomarker, and we therefore reviewed published values of normal variation MTR. Larger values may indicate a greater contribution from instrumental variation (assuming the intrinsic biological variation does not vary between centres).

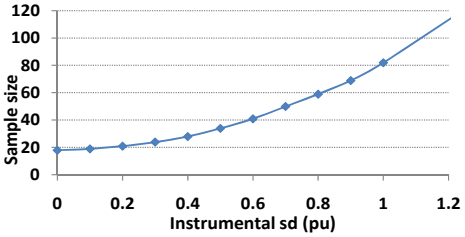

Figure 1: Sample size increases with instrumental sd.

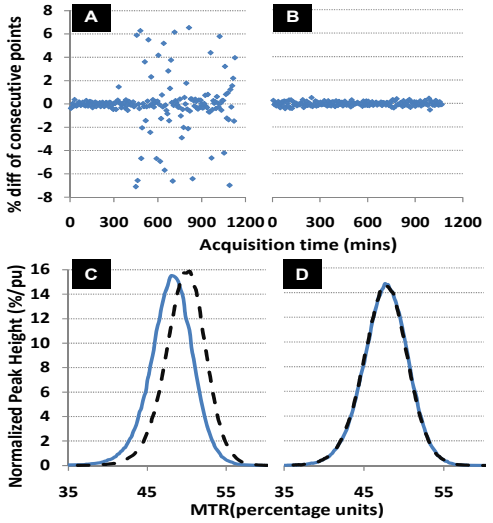

Figure 2: Phantom measures (A and B) and MTR histograms (C and D) before (left) and after (right) changing the transmitter boards.

**Results:** *1. Modelling dependence of statistical power on instrumental variance.* Depending on instrumental sd, the required sample size varied from 18 to 82 (fig 1).
*2. Measurement of normal range*. The initial normal sd was 0.92 (95% CL 0.6-1.7); after board replacement this decreased to 0.39 (95% CL 0.26-0.75) ("before" and "after" on fig 3)
*3. Instrumental variation in controls* was sd=0.67(95%CL 0.47-1.32), and histograms had shifts up to 1.76pu (fig 2c). After replacement of MRI transmitter boards instrumental variation decreased to sd=0.12 (95%CL 0.07-0.45) ("ISD" on fig 3), and individual MTR histograms appeared identical (fig 2d).
*4. Instrumental variation in a phantom* showed signal values having a range of 8% (fig 2a); the sd of the running difference was 2.1%. After replacement of the transmitter boards (fig 2b) the running difference was down to 0.2%.
5. *Published values of normal variation* for MTR in white matter in healthy controls ranged from 0.4 to 1 sd (fig 3) (mean values are 30-48 pu). Following the replacement of defective hardware our sd dropped from the high end to the low end of the published range.

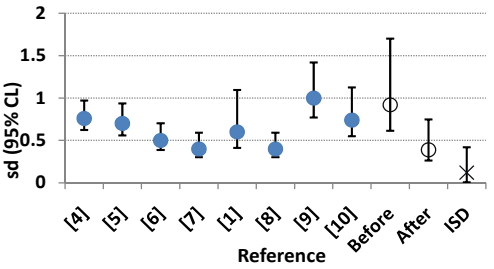

Figure 3: Normal variation (sd) for white matter MTR. Published values [1,4-10], measured values before and after changing the transmitter boards, and instrumental variation - ISD (after only).

**Future work: 1**. Instant repeatability measurements will continue as part of ongoing QA.
**2.** Monthly repeatability measurements will be made. **3.** Correction for the effect of age would give group sd estimates that are more comparable.

**Discussion and conclusions: 1.** Ongoing QA for quantitative studies should include explicit measurement of short- and long-term repeatability in controls. Although CV's of a few percent might indicate good performance, group differences are also only a few percent.
**2.** Instrumental variation can have a crucial influence on the sensitivity of quantitative brain measurements and their statistical power to detect subtle effects. **3.** Ongoing QA should also demonstrate that measured normal variation is at the low end of the published range. **4**. Long overnight phantom scanning sessions can demonstrate machine stability or the lack of it.

**References: 1.** Tofts, P.S., et al., Magma, 2006. 19(4): p. 209-22. **2.** Steens, S.C., et al., Eur Radiol, 2004. 14(3): p. 425-30. **3.** Howell, d., *Statistical Methods for Psychology - 4th edition*. 1997. P 205-06 **4.** Silver, N.C., et al., J Neurol Neurosurg Psychiatry, 1997. 62(3): p. 223-28. **5.** Ge, Y.L., et al., Am J Neuroradiol, 2002. 23(8): p. 1334-41. **6.** Steens, S.C., et al., Arthritis Rheum, 2004. 50(9): p. 2877-81. **7.** Davies, G.R., et al., Journal of Neurology, 2005. 252(9): p. 1037-44. **8.** Khaleeli, Z., et al., J Neurol Neurosurg Psychiatry, 2007. 78(10): p. 1076-82. **9.** Vrenken, H., et al., Multiple Sclerosis, 2007. 13(6): p. 708-16. **10.** Bellmann-Strobl, J., et al., Eur Radiol, 2009. 19(8): p. 2066-74.