

Do DTI reproducibility studies agree? A meta-analysis.

Pim Pullens¹, Wim van Hecke², and Paul M Parizel¹

¹Radiology, Antwerp University Hospital, Edegem, Antwerp, Belgium, ²Icometrix, Leuven, Belgium

Audience clinicians, neuroscientists, clinical trial researchers

Purpose Since DTI is increasingly used in clinical trials and routine clinical care, it is of paramount importance to assess normal variation and reproducibility of DTI based measures. A number of studies have recently investigated reproducibility, using different scanners, protocols and analysis methods. Notwithstanding these efforts, a review of these results is needed to evaluate agreement, which could ultimately lead to acceptance of DTI measures as a biomarker. A common method to assess agreement is to look at coefficients of variation (CV), but it is often overlooked that comparison of CVs is not straightforward [1,2], especially when sample sizes differ. We performed a meta-analysis using appropriate hypothesis tests [3] on FA values in the corpus callosum across a number of published reproducibility studies and investigated the effect of field strength, inter/intra-rater reproducibility and scanner type.

Methods A non-exhaustive search was performed on Pubmed and Web of Knowledge using any combination of keywords "DTI, Diffusion Tensor Imaging, DWI, Diffusion Weighted Imaging, reproducibility, reliability multi-center, quality assurance, QA". Out of 28 search results, papers were selected that included healthy adult subjects, and that report tabular data on mean μ and standard deviation σ of FA values in the corpus callosum. All papers report ROI measures, except for #3 (fiber tracking) and #11 (automated image processing). Based on the selection criteria, 6 papers were included in this study. We base our comparison on $CV = \sigma/\mu$. We used the methods described in [2], which in short, allows to perform a test of the hypothesis that CVs associated with k populations, each with N samples are all equal, available at <http://www1.fpl.fs.fed.us/covtestk.html>. Tabular data was extracted automatically from the PDFs using Tabula (<http://tabula.nerdpower.org/>). FA σ , μ in the splenium of the corpus callosum were extracted from the papers. 95% confidence intervals on the mean were calculated according to [3].

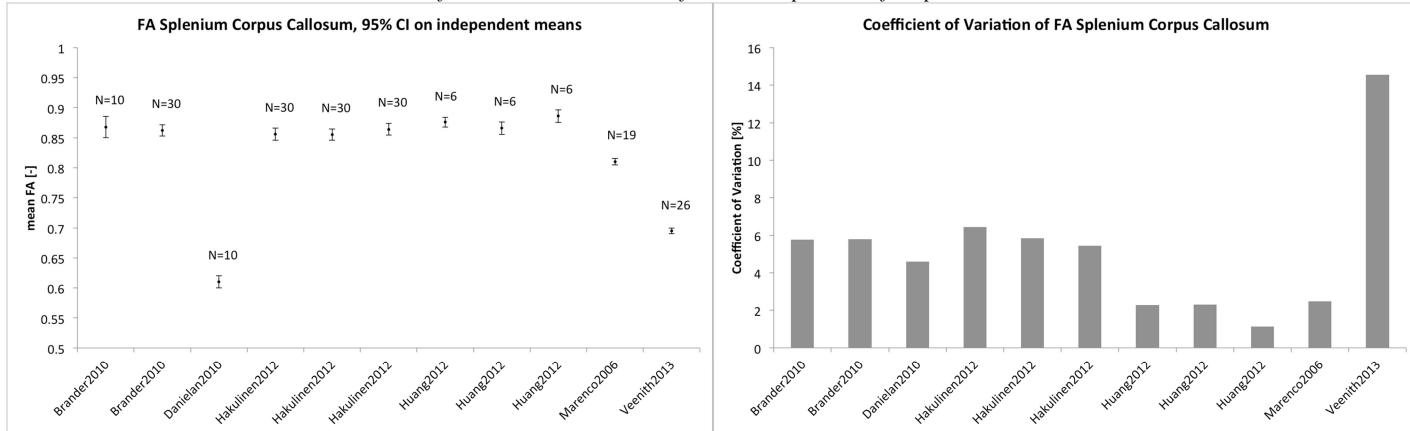
Results

Description of included studies. N=number of subjects. S=nr of scans, B=field strength, b=b-value [mm²/s].

Paper	N	Age	S	B	scanner	Coil	#dirs	b	Matrix	FOV	gap	recon size	TR	TE	NEX
1 Brander2010[4]	10	38.3(19-61)	1	1.5	Siemens Avanto	12 ch head	12	1000	128x128	230	1.5	1.8x1.8x5	3600	96	3
2 Brander2010	30	38.3(19-61)	1	3	Siemens Trio	12 ch head	20	1000	128x128	230	0.9	1.8x1.8x3	5144	92	3
3 Danielan2010[5]	10	59.3(50-72)	1	3	Philips Intera	8 ch head	32+1	1000	96x96	240	0	1.875x1.875x2.5	N/A	86	4
4 Hakulinen2012[6]	30	37.8(18-60)	1	3	Siemens Trio	12 ch head	20	1000	128x128	230	0.9	1.8x1.8x3	5144	92	3
5 Hakulinen2012	30	37.8(18-60)	1	3	Siemens Trio	12 ch head	20	1000	128x128	230	0.9	1.8x1.8x3	5144	92	3
6 Hakulinen2012	30	37.8(18-60)	1	3	Siemens Trio	12 ch head	20	1000	128x128	230	0.9	1.8x1.8x3	5144	92	3
7 Huang2012[7]	6	26+/4	1	3	Siemens Trio	32 ch head	60+8	1000	112x130	N/A	0	2x2x2	9000	83	1
8 Huang2012	6	26+/4	1	3	Siemens Trio	32 ch head	60+8	1000	112x130	N/A	0	2x2x2	9000	83	1
9 Huang2012	6	26+/4	1	3	Siemens Trio	32 ch head	60+8	1000	112x130	N/A	0	2x2x2	9000	83	1
10 Marenco2006[8]	19	26+/4.4(21-36)	2	1.5	GE Signa	N/A	6+1	1100	N/A	N/A	0	2x2x2	>10000	82.7	8
11 Veenith2013[9]	26	34(25-44)	2	3	Siemens Verio	N/A	63+1	1000	192x92	N/A	0	2x2x2	11700	106	1

Notes #3: fiber tracking #4: observer 1, measurement 1, #5 obs 1, meas 2 #6 obs 2, meas 2 #7: scanner 1, #8: scanner 1 after coil replacement #9: second identical scanner as #8.

Confidence intervals and CV of FA in the splenium of corpus callosum



Hypothesis tests of CVs

Case	Accept?	p-value
Are CVs across all studies equal?	Yes	0.4173E-19
What if we remove #3, fiber tracking?	Yes	0.5580E-19
Are CVs across 3T, ROI based studies equal?	Yes	0.1058E-14
Are intra- or inter-rater CVs equal? Intra-rater (4 = 5)	No	0.6077
Inter-rater (5 = 6)	No	0.6929
Are CVs from 2 identical scanners equal (8 = 9)?	No	0.9178E-01
Is 1.5T equal to 3T (1 = 2)?	No	0.8701

found no agreement, which is likely to be related to the small amount of data. We propose that reliability studies should report tabular data in order to find agreement. **In conclusion**, CVs of 3T DTI measures are comparable independent of acquisition found in these studies, but individual subject's data should be processed similarly, preferable with manual ROI analysis, to make a fair comparison and identify confounds in reproducibility studies.

1.Lewontin, *Systematic Zoology* 1966, 15:2:141-142; 2. Verrill, *Communications in Statistics—Theory and Methods* 2007, 36:2187-2206; 3. www.cochrane-handbook.org version 5.0.0; 4.Brander *Acta Radiol* 2010, 51:800; 5.Danielan *NeuroImage* 2010, 49:1572-1580; 6.Hakulinen *BMC Medical Imaging* 2012, 12:30; 7.Huang, *Plos One* 2012, 7:10 e47684; 8.Marenco *Psychiatry Research: Neuroimaging* 2006,147:69-78; 9.Veenith, *Plos One* 2013, 8:6:e65941

Discussion Although this overview of DTI reproducibility studies is limited due to the variability of acquisition and analysis approaches across these studies, some interesting observations can be made: CIs on the mean FA show that study #3, #10 and #11 deviate from the rest of the data. #3 used fiber tracking, #10 a 1.5T GE scanner, and #11 automated image processing, while other papers used manual ROI analysis. Based on CVs, results across all studies seem comparable. In the reviewed cases we