NeuroImage: Clinical 15 (2017) 601–612

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

In utero diffusion tensor imaging of the fetal brain: A reproducibility study

András Jakab[a,b,*], Ruth Tuura[a], Christian Kellenberger[c], Ianina Scheer[c]

[a] Center for MR-Research, University Children's Hospital, Zürich, Switzerland
[b] Computational Imaging Research Lab (CIR), Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria
[c] Department of Diagnostic Imaging, University Children's Hospital, Zürich, Switzerland

ARTICLE INFO

Keywords:
Fetal diffusion MRI
Diffusion tensor imaging
Fetal brain connectivity
Prenatal development
Connectome

ABSTRACT

Our purpose was to evaluate the within-subject reproducibility of *in utero* diffusion tensor imaging (DTI) metrics and the visibility of major white matter structures.

Images for 30 fetuses (20–33. postmenstrual weeks, normal neurodevelopment: 6 cases, cerebral pathology: 24 cases) were acquired on 1.5 T or 3.0 T MRI. DTI with 15 diffusion-weighting directions was repeated three times for each case, TR/TE: 2200/63 ms, voxel size: 1 * 1 mm, slice thickness: 3–5 mm, b-factor: 700 s/mm^2. Reproducibility was evaluated from structure detectability, variability of DTI measures using the coefficient of variation (CV), image correlation and structural similarity across repeated scans for six selected structures. The effect of age, scanner type, presence of pathology was determined using Wilcoxon rank sum test.

White matter structures were detectable in the following percentage of fetuses in at least two of the three repeated scans: corpus callosum genu 76%, splenium 64%, internal capsule, posterior limb 60%, brainstem fibers 40% and temporooccipital association pathways 60%. The mean CV of DTI metrics ranged between 3% and 14.6% and we measured higher reproducibility in fetuses with normal brain development. Head motion was negatively correlated with reproducibility, this effect was partially ameliorated by motion-correction algorithm using image registration. Structures on 3.0 T had higher variability both with- and without motion correction.

Fetal DTI is reproducible for projection and commissural bundles during mid-gestation, however, in 16–30% of the cases, data were corrupted by artifacts, resulting in impaired detection of white matter structures. To achieve robust results for the quantitative analysis of diffusivity and anisotropy values, fetal-specific image processing is recommended and repeated DTI is needed to ensure the detectability of fiber pathways.

1. Introduction

Since the first depiction of the diffusion process in the human brain significant conceptual and methodological developments have been applied to diffusion MRI (Le Bihan and Johansen-Berg, 2012), leading to the widespread use of various MRI techniques based on this phenomenon, such as diffusion-weighted imaging, diffusion tensor imaging (DTI, (Basser et al., 1994; Westin et al., 2002)), intravascular incoherent motion (IVIM) imaging (Le Bihan et al., 1988), diffusion kurtosis imaging (Bar-Shir et al., 2009; Raab et al., 2010) or diffusion spectrum imaging (DSI, (Sotiropoulos et al., 2013)). DTI offers increased sophistication over diffusion-weighted MRI since it provides information about both the magnitude and orientation of the anisotropic diffusion in tissues, consequently allowing the calculation of the magnitude of diffusion anisotropy, and the parallel and perpendicular diffusivity. According to basic experiments, such parameters in the human brain reflect the underlying axonal membrane microstructure (Beaulieu, 2002), correlate with myelination patterns (Dubois et al., 2013), can serve as group-level markers of various brain pathologies (Le Bihan and Johansen-Berg, 2012; Basser and Pierpaoli, 1996; Arfanakis et al., 2002; Lu et al., 2003; Tropine et al., 2004; Price, 2007; Hess, 2009; Babikian et al., 2009), and allow more complex image post processing approaches, such as fiber tractography (Basser et al., 2000; Catani et al., 2002). Diffusion MRI approaches are therefore widely regarded as nascent methods for characterising the human brain from its tissue microstructure to its network-level connectional architecture, referred to as the connectome (Sporns, 2011). The usability of DTI, owing to recent advancements in fast imaging sequence development, can be extended to the earliest point of the human lifespan: before birth, as early as the second trimester of gestation (Kasprian et al., 2010). Initial *in utero* DTI studies have revealed how commissural, projection and association fibers emerge in the living human fetus

Abbreviations: AD, axial diffusivity; CCA, corpus callosum agenesis; CV, coefficient of variation; DTI, diffusion tensor imaging; FA, fractional anisotropy; GW, gestational week; MD, mean diffusivity; RD, radial diffusivity; ROI, region of interest; SSIM, structural similarity index

* Corresponding author at: Center for MR-Research, University Children's Hospital, Steinwiesstrasse 75, CH-8032 Zürich, Switzerland.
E-mail address: andras.jakab@kispi.uzh.ch (A. Jakab).

http://dx.doi.org/10.1016/j.nicl.2017.06.013
Received 28 November 2016; Received in revised form 25 January 2017; Accepted 8 June 2017
Available online 09 June 2017

(Kasprian et al., 2008; Mitter et al., 2015a), detected pathological, ectopic fibers (Kasprian et al., 2013; Mitter et al., 2015b; Jakab et al., 2015), and shown surprisingly good agreement with similar, *post mortem* fetal MRI studies (Huang et al., 2009).

As DTI was applied to ever more demanding experimental and clinical scenarios, it faced numerous tests of reproducibility, resulting from the increased sensitivity of DTI and other echo-planar imaging (EPI) sequences towards artifacts. Such experiments are of key importance when we consider the so-called secondary maps – such as the fractional anisotropy – quantitative makers of tissue microstructure in normal and pathological conditions. Although DTI, especially in adults and children, can be used to calculate the scalar metrics of diffusion with high confidence (Pfefferbaum et al., 2003; Heiervang et al., 2006; Bonekamp et al., 2007; Farrell et al., 2007), and tractography has also been shown to be reproducible under constant conditions (Wakana et al., 2007; Malykhin et al., 2008; Danielian et al., 2010; Besseling et al., 2012; Owen et al., 2013), it is now clear that differences in the scanner type or imaging site (Cercignani et al., 2003; Landman et al., 2007; Sasaki et al., 2008; Fox et al., 2012; Jakab et al., 2016), the parameter settings of the sequence and possibly other factors greatly hinder the comparability across research sites. Fetal DTI is further complicated by excessive motion and more pronounced susceptibility artifacts because of the heterogeneous chemical composition of the maternal organs surrounding the fetal head.

Despite the technical challenges associated with *in utero* DTI, it is currently the only clinically viable imaging modality capable of visualizing the developing white matter during the second and third trimesters of gestation. The diffusion tensor approach provides added value to that of diffusion-weighted imaging in that the fractional anisotropy, the eigenvalues and orientations of the tensor may reflect many, thus-far not thoroughly studied, physiological and microstructural attributes of fetal nerve tissue. DTI and tractography before birth therefore represent important approaches for basic and clinical neuroscience research, especially since our current understanding of the transient morphology of the emerging human brain pathways and its vulnerability during “risk periods” of development are based on postmortem histology and imaging in a limited number of subjects. *In vivo* MRI validations of advanced early brain imaging data are only available from the 24-26th weeks in extremely preterm neonates (Doria et al., 2010), which is a suboptimal model for the characterization of white matter development. In contrast to the *post mortem*, animal investigations and postnatal data of preterm infants, a prenatal DTI based work-up augmented with tractography can provide new insight into fiber development in normally developing fetuses or reveal the *in utero* trajectory of pathological fiber development. Such a rich parameter set may not only be viewed as a fingerprint of normal and pathological development on the case-level basis, but as potential candidates for prenatal biomarkers, and thus they can hold the clue to predicting the outcome of pregnancy and the neurological outcome after birth.

This endeavor, however, should be preceded by validation experiments aiming to characterise the reliability of the quantitative metrics of diffusion, and to identify the possible factors that can confound their applicability as markers of disease in fetal MRI. Our purpose, therefore, was to demonstrate the within-subject reproducibility of *in utero* DTI in a clinical cohort of fetuses with unaffected and pathological brain development taken from a clinical sample, and to evaluate how scanner field strength, fetal age, fetal motion patterns or the presence of pathology affect the reproducibility of the DTI derived metrics of diffusion.

2. Methods

2.1. Study population

The study population consisted of fetuses for which fetal MRI was clinically indicated. The general clinical indication for fetal MRI is in conditions where the combined accuracy of MRI and ultrasound is higher than with the ultrasound alone. In our study, the indication to

Table 1
Demographic data of the study population.

Subject ID	MRI scanner field strength	Gestational age (week + day)	Main neuroimaging findings
1	3.0 T	24 + 5	Myelomeningocele and Chiari II malformation
2	1.5 T	23 + 2	Ventricular dilatation, thin tectum, subinsular T2 hyperintensity
3	1.5 T	27 + 1	Ventricular dilatation, thin tectum, subinsular T2 hyperintensity
4	3.0 T	30 + 2	Total agenesis of the corpus callosum
5[a]	1.5 T	28 + 3	Moderate dilatation of the lateral ventricles, resorption of the septum pelludicum
6[a]	1.5 T	28 + 3	*Normal brain morphology*
7	1.5 T	23 + 0	*Normal brain morphology*
8	3.0 T	30 + 4	Total agenesis of the corpus callosum
9	3.0 T	24 + 0	Microcephaly
10	3.0 T	22 + 6	Mild dilatation of the lateral ventricles
11	1.5 T	31 + 0	Total agenesis of the corpus callosum
12	3.0 T	20 + 0	*Normal brain morphology*
13	1.5 T	31 + 4	Asymmetric dilatation of the lateral ventricles, septation of the ventricles
14	1.5 T	29 + 3	Mild dilatation of the lateral ventricles
15	1.5 T	28 + 1	Postoperative status of fetal correction for myelomeningocele
16	1.5 T	24 + 0	*Normal brain morphology*, *thorax abnormality*
17	1.5 T	25 + 0	*Normal brain morphology*
18	3.0 T	28 + 1	Arachnoid cyst in posterior fossa
19	3.0 T	24 + 4	Spina bifida and Chiari II malformation
20	1.5 T	22 + 5	Spina bifida and Chiari II malformation
21	3.0 T	30 + 4	Mild dilatation of the lateral ventricles
22	1.5 T	22 + 6	*Normal brain morphology*
23	1.5 T	30 + 4	Mild dilatation of the lateral ventricles
24	3.0 T	28 + 4	Mild dilatation of the lateral ventricles
25	3.0 T	33 + 0	Enlargement of the cisterna magna
26	3.0 T	32 + 0	Vena Galeni malformation
27	3.0 T	26 + 0	Total agenesis of the corpus callosum
28	1.5 T	26 + 0	Total agenesis of the corpus callosum
29	1.5 T	26 + 3	Spina bifida and Chiari II malformation
30	1.5 T	27 + 0	Spina bifida and Chiari II malformation

[a] Cases 5 and 6 are dizygotic twins imaged at the same time with separate DTI acquisitions.

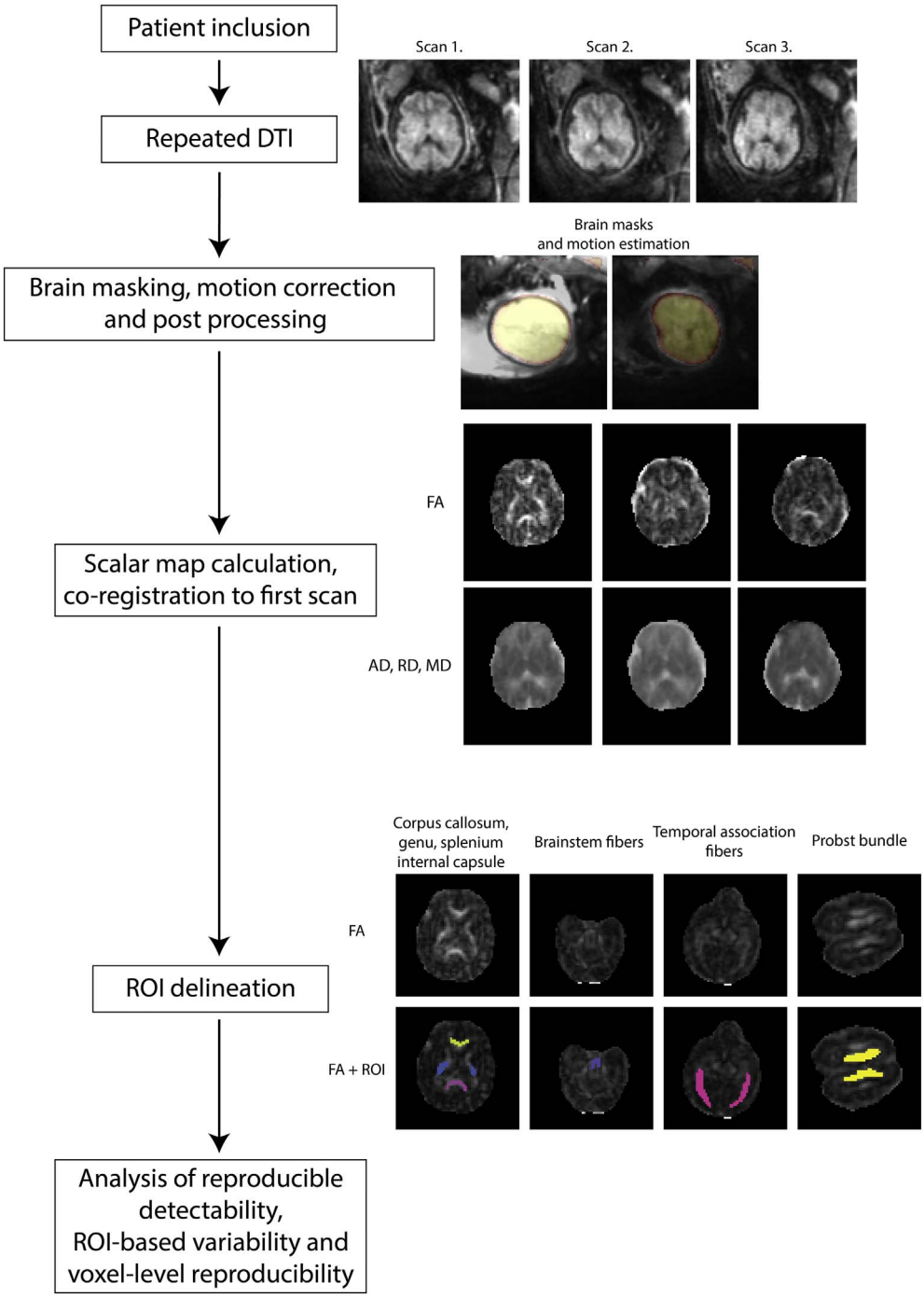

Fig. 1. Testing the reproducibility of *in utero* DTI: overview of the study work-flow.

perform fetal MRI was to rule out or confirm brain or lung pathologies or for post-operative follow-up after open fetal correction for spina bifida. Demographic data, clinical indication and the summary of neuroimaging findings are summarized in Table 1. As part of the routine clinical protocol, fetal MRI included (Le Bihan and Johansen-Berg, 2012) multi-planar structural MRI examinations with T2-weighted sequences of the whole fetus, fetal brain and the placenta, (Basser et al., 1994) T1-weighted and echo planar diffusion-weighted sequences to rule out intracranial hemorrhage and/or blood breakdown products, and (Westin et al., 2002) DTI, for which the non-processed, trace-weighted *isotropic image* was mostly used in clinical diagnostics.

Inclusion criteria were: fetal age equal to or higher than 20 weeks of gestation based on ultrasound report prior to the MRI, availability of three repeated-session DTI scans with only a few motion-corrupted time frames (< 5). During the study period between January 2016 and January 2017, we enrolled 30 fetuses in the study, with a mean gestational age of 27 $\pm$ 3.3 (range: 20–33) weeks. Of 30 fetuses, 6 had normal brain development according to the ultrasonography and fetal MRI reports. The mothers gave written, informed consent for use of their clinical data for research purposes prior to the examination, and the research was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the regional ethical committee.

2.2. Fetal diffusion tensor imaging protocol

As a part of clinical routine, fetal MRI was performed on two different clinical MRI systems, one with a field strength of 1.5 T and one of

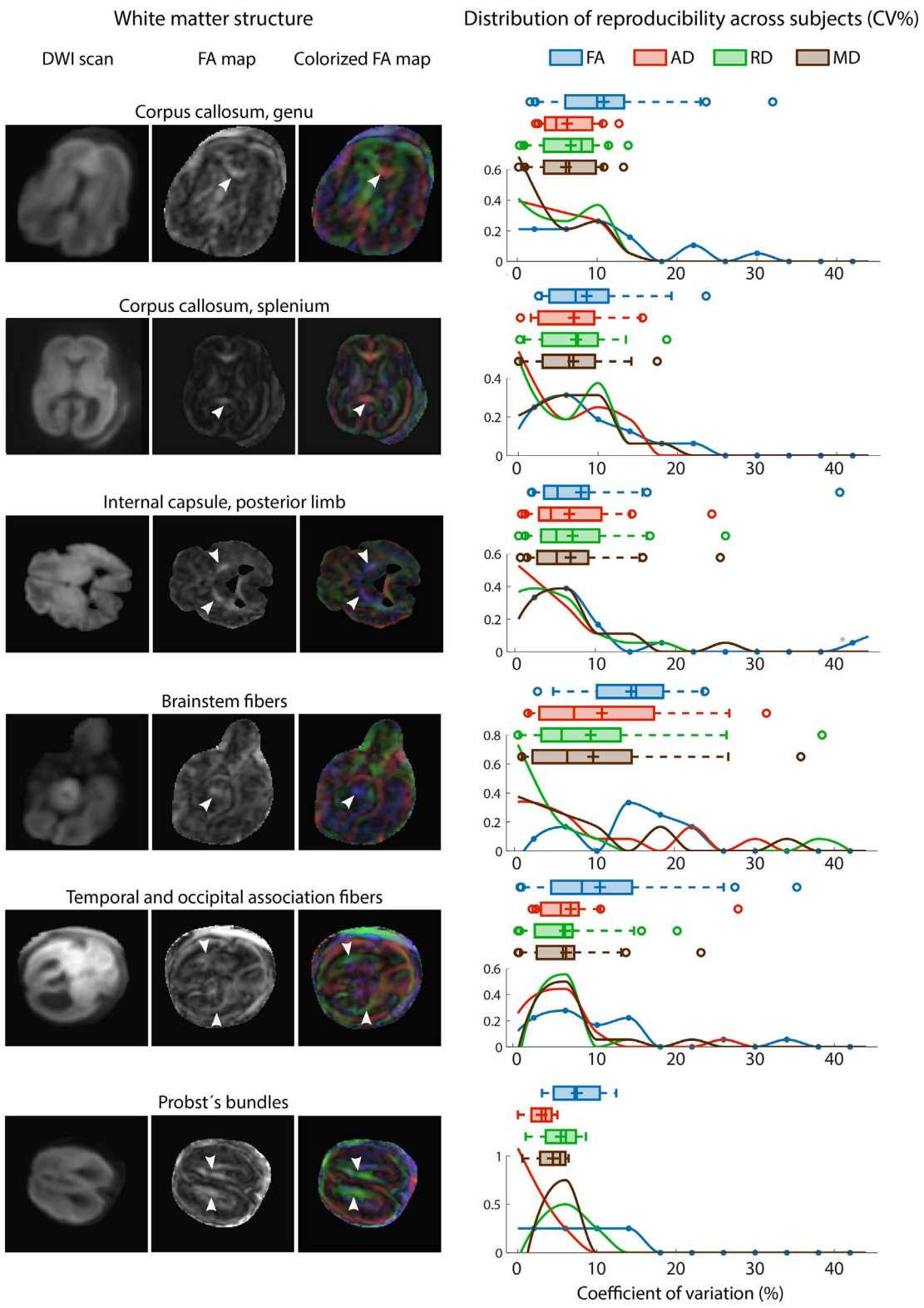

Fig. 2. Reproducibility of DTI derived metrics in selected white matter structures. We illustrate each of the evaluated white matter structures on the mean (trace-weighted) diffusion image, on the fractional anisotropy and on the colorized fractional anisotropy maps (left column). The reproducibility of the fractional anisotropy (FA), axial diffusivity (AD), radial diffusivity (RD) and the mean diffusivity (MD) is given as the coefficient of variability (CV%) across the repeated scans. The distribution of CV% across the study population for each white matter structure and for each evaluated DTI metric is graphed separately (right column). In each diagram, the Whisker plots represent the interquartile range (bar), mean value (horizontal line), median value (cross), outliers (circles), while the histogram demonstrates the distribution of CV% values for each DTI metric separately.

3.0 T (MR450 and MR750, GE Healthcare, Milwaukee, WI, USA). The assignment of the cases to either of these scanners was not controlled in the current study, and was based on the availability of free scanner time. Pregnant women were examined in the supine position (feet first), and no contrast agents or sedatives were administered. In order to obtain optimal MR signal, the coil was readjusted depending on the position of the fetal head during the imaging procedure. DTI scans followed the structural T2-weighted images, and three repeated DTI

sessions were performed consecutively or with other imaging sequences in-between. Axial slices were positioned orthogonal to the fetal brainstem. The basic settings of the DTI sequence were identical for the two MR scanners used. For DTI acquisitions, an axial, single-shot, echo planar imaging sequence was used with a TR of 2200 ms, a TE of 63 ms, acquisition matrix of 112 ∗ 112 re-sampled to 256 ∗ 256, a voxel size of 1 ∗ 1 mm, and a slice thickness 3 to 5 mm without a gap or interleaved slices. Depending on the size of the fetal brain and gestational age, 10–18 slices were acquired, covering the whole brain from the brainstem to the convexities. Images were acquired using SENSE and a pseudo-receive bandwidth of 33 Hz/pixel along the phase encode dimension. For each of the three repeated DTI scans, 15 non-collinear diffusion-weighted magnetic pulsed gradients were used with a b value of 700 s/mm^2 and one B_0 image without diffusion weighting was also collected. Fetal DTI data were anonymized and transferred to image-processing workstations in NIfTI image format.

2.3. Standard and fetal-specific post processing of DTI data

After transferring the MRI data to the image processing workstations the following analysis steps were performed: (Le Bihan and Johansen-Berg, 2012) manual masking of the fetal brain volume to separate non-brain tissue, (Basser et al., 1994) fetal-specific post processing including motion correction of DTI frames, (Westin et al., 2002) standard diffusion tensor estimation and calculation of the scalar measures of diffusion, (Le Bihan et al., 1988) detection of white matter structures and region of interest (ROI) placement on major fiber bundles using colorized fractional anisotropy and fractional anisotropy images, (Bar-Shir et al., 2009) qualitative and quantitative analysis of image reproducibility.

Due to the presumed confounding effects of fetal head and maternal respiratory motion on the quality of the reconstructed DTI data, we corrected the images for spurious fetal movements. First, the fetal brain was manually masked on the first image frame of each of DTI scan. To ensure that the brain's borders remain precisely within the mask despite the possible movements in the consecutive images, a machine learning algorithm was utilized that propagated this mask along the time dimension in each scan. This was followed by the co-registration of each masked image frame to the first reference image of the scan to achieve identical orientation and good anatomical overlap. The fetal-specific image processing algorithm is described in detail in the Supplementary document. After the re-orientation of the raw diffusion-weighted images to reduce the effects of fetal head motion, we performed standard diffusion tensor estimation using the *dtifit* command in the FDT toolkit of the FSL software package. For the diffusion estimation, the b-vector orientations were corrected for the estimated head rotation, and a weighted least squares approach was used. We calculated the fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD) from the tensor datasets using the known general equations from the literature (Basser and Pierpaoli, 1996). For further post-processing steps, the mean image of all the non-B_0 image frames was calculated for each scan; we will refer to this image as the “mean DWI”.

The next steps of the analysis included the qualitative assessment of fiber visibility on each repeated scans, ROI placement and the voxelwise comparison of values across repeated scans. For the last two steps, good anatomical overlap between the repeated imaging within each subject was required. Based on the masked images of the mean DWI, two transformation matrices were determined that transformed the second and third repeated scans to the space of the first DTI of each subject. This linear co-registration was performed by the *flirt* command in the FSL software package, which utilized the least square differences as the cost function and optimized the transformation using 6 degrees of freedom. These matrices were used to re-sample the repeated FA, MD and AD scalar maps to enable the comparison of voxelwise values and ROI-based values. An overview of our study work-flow is illustrated in Fig. 1.

2.4. Definition of white matter structures

Due to the ongoing development of white matter structures during mid-gestation and the technical difficulties of imaging, we selected five prenatally visible major fiber bundles of normal brain development and evaluated whether they can be detected reproducibly across repeated DTI scans. Furthermore, the Probst's bundles were also evaluated in cases with total agenesis of the corpus callosum. A structure was classified detectable if a physician with experience in fetal anatomy could see it on at least two axial slices of the fractional anisotropy images. We evaluated the (Le Bihan and Johansen-Berg, 2012) anterior and (Basser et al., 1994) posterior part of the corpus callosum (genu and splenium, respectively) at the level of the third ventricle, (Westin et al., 2002) the bilateral internal capsule, posterior limb, (Le Bihan et al., 1988) the brainstem fibers consisting of projection fibers including the corticospinal tract and the medial lemnisci at the level of the pons to mesencephalon and the (Bar-Shir et al., 2009) fiber pathways that surround the occipital horn of the lateral ventricles posterolaterally. This latter structure comprised the inferior fronto-occipital fascicle, the optic and acoustic radiations and the subcortical fibers of the temporal and occipital lobe, and therefore we used the term “temporooccipital association fibers” in our study. In 3 cases with corpus callosum agenesis, the ectopic bundles of Probst were evaluated instead of the callosal fibers. These ROIs were saved for each repeated scan, and the underlying DTI derived metrics were averaged over the ROI using the FSL software package. Each white matter structure is depicted in Fig. 2.

2.5. Analysis of white matter structure detectability

In our study reproducibility was defined using two concepts: the ability to detect major white matter structures in repeated scans within one MRI session in each individual and the variability of quantitative measurements of a given structure across the repeated scans. We tested the reproducibility of *in utero* DTI using the following qualitative and quantitative approaches: (Le Bihan and Johansen-Berg, 2012) the observer based visibility of white matter structures on FA maps, (Basser et al., 1994) reproducibility of DTI derived measures in anatomically important regions of interests (ROIs) and the (Westin et al., 2002) voxelwise variability, image correlation and image similarity measured over the whole fetal brain. During the qualitative analysis of tract visibility and reproducibility, the following categories were assigned to each tract in each subject. A tract was classified “not visible”, if it cannot be seen on any of the repeated scans, “visible, not reproducible” if it was only seen on one scan, “moderately reproducible” if it was visible on two scans, “highly reproducible” if it was visible on all three scans. For each subject, the occurrence was also calculated, which captured how often a particular tract is visible during the repeated scans (%).

2.6. Analysis of repeatability

After this qualitative evaluation, the FA, MD, RD and AD values were measured in each ROI where the tract was detectable on at least two of the FA images. Variability was quantified as the coefficient of variation (CV) in percent units across repeated scans using the following equation:

$$Variability = CV = \frac{\sigma}{\overline{x}} \tag{1}$$

where σ is the standard deviation across the repeated scans, and $\overline{x}$ is the mean value across the repeated scans. σ is defined as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{2}$$

where N is the number of repetitions in which the tract is visible, $\overline{x}$ is

Table 2
Visibility of major white matter structures *in utero* across the repeated DTI scans. A structure was classified "not visible", if it cannot be seen on any of the repeated scans, "not reproducible" if it was only seen on one scan, "moderately reproducible" if it was visible on two of the three scans, "highly reproducible" if it was visible on all three scans. Frequency means the average detectability of the tract in an individual across the repeated scans (0–100%).

Name of structure	Not visible, % (number of subjects)	Not reproducible, % (number of subjects)	Moderately reproducible, % (number of subjects)	Highly reproducible, % (number of subjects)	Frequency (%, mean ± SD, min-max)
Corpus callosum, genu	16% (4/25)	8% (2/25)	28% (7/25)	48% (12/25)	69.3 ± 37.1%
Corpus callosum, splenium	20% (5/25)	16% (4/25)	28% (7/25)	36% (9/25)	60 ± 38.5%
Internal capsule, posterior limb	6.67% (2/30)	33.3% (10/30)	20% (6/30)	40% (12/30)	64.4 ± 33.8%
Brainstem fibers	30% (9/30)	30% (9/30)	26.7% (8/30)	13.3% (4/30)	41.1 ± 34.7%
Temporal and occipital association fibers	20% (6/30)	20% (6/30)	16.7% (5/30)	43.3% (13/30)	61.1 ± 40.2%
Probst's bundles	0% (0/5)	20% (1/5)	20% (1/5)	60% (3/5)	66.7 ± 33.3%

the mean value across the repeated scans, x_i is the ith value.

The voxelwise reproducibility of images tested whether values across the entire brain could be reproduced. For this analysis step, the images were co-registered to the first scan of each subject in order to achieve anatomical correspondence. A Gaussian filter was applied to the images with a full width at half maximum (FWHM) of 2 mm to correct for small remaining sampling errors due to the imperfect overlap between the scans.

Three parameters were used to characterise the reproducibility of each DTI metric on a per-voxel basis. First, we calculated the standard deviation of DTI metrics using Eq. (2) and then averaged the values over all brain voxels. The image correlation was calculated using the Pearson product-moment correlation coefficient, averaged over all brain voxels. We then calculated the Structural Similarity (SSIM) Index for each image pair (Wang et al., 2004). SSIM is a quality assessment index, which is based on the computation of three terms, namely the luminance term, the contrast term and the structural term. The overall index is a multiplicative combination of the three terms:

$$SSIM(x,y) = [l(x,y)]^{\alpha*}[c(x,y)]^{\beta*}[s(x,y)]^{\gamma} \quad (3)$$

where for a given pixel x, y, l, c and s are the luminance term, the contrast term and the structural term, and

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, \quad (4)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \quad (5)$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \quad (6)$$

where μ_x, μ_y, σ_x, σ_y, and σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y.

We evaluated whether the gestational age of the fetus, the field strength of the scanner, the applied motion correction, the magnitude of fetal head movement and the position of the fetal head influenced the reproducibility measures. The influence of these factors on the white matter structure detectability was tested by comparing the reproducibility measures between the groups formed by the categorical variables, such as scanner field strength, using the Wilcoxon rank sum test. The influence of continuous variables, such as the gestational age and fetal head movement on reproducibility was tested by linear regression, and outlier effects were addressed using the least squares fitting approach in Matlab for Windows R2014 (Mathworks Inc., Natick, MA, USA). For this analysis, we report the coefficient of determination (R-squared).

3. Results

3.1. Visibility of white matter structures across repeated scans

The corpus callosum was investigated in 25 cases; in the remaining 5 cases, due to the total agenesis of the corpus callosum, ectopic bundles were examined in place of the corpus callosum. The genu and splenium of the corpus callosum were the most commonly visible structures in our study, demonstrating a moderate to high reproducibility in 76% (genu) and 64% (splenium) of the cases, meaning that the structures were detectable in at least 2 of the 3 repeated DTI scans. However, the central part of the corpus callosum was usually not visible, the only exceptions were larger fetuses in whom the thicker corpus callosum was generally easier to delineate. Without motion correction, the visibility of these structures was considerably lower: 60% and 44%, respectively. The bilateral internal capsule's posterior limbs were successfully visualized in 93.3% of the cases in at least one of the repeated DTI scans, although in 33.3% of the fetuses this structure was not reproducible and was only seen in one of the three scans, while in 40% of the cases it was seen in all of the three repeated scans. Without correcting the images for fetal motion, the internal capsule, posterior limb was highly reproducible only in 20% of the cases, and was not visualized in 9 cases. The temporooccipital association fibers, which include different pathways lateral to the ventricles in the temporal and occipital lobe, were moderately to highly reproducible (moderate reproducibility: 16.7%, high reproducibility: 43.3% of the cases). The least detectable structures in our analysis were the brainstem fibers, which were only seen repeatedly in 40% of the cases and high reproducibility was only reported for 4 fetuses (13.3%). The ectopic Probst bundles were seen repeatedly in 4 of the 5 CCA cases. The qualitative visibility results after fetal-specific image processing for each structure are summarized in Table 2, while the non-motion corrected values are given in Table S1.

3.2. Reproducibility of DTI measures

The results of the reproducibility analysis for each white matter structure are summarized in Fig. 2. The variability of DTI measures across repeated scans in the individual fetuses spanned a large interval from 0.2% to 40.4% with several outliers, *e.g.* individual cases demonstrating a high variability for all tracts. The variability of values for a given structure averaged over the population was between 3% (Variability of axial diffusivity of the Probst bundle) and 14.6% (Variability of FA, brainstem). Generally, the mean diffusivity, which is rotationally invariant, was less variable, while the FA was the most variable of the examined DTI metrics. The pathological Probst bundle showed considerably lower variability (3% - 7.5%) compared to the other structures. The tracts with the lowest variability after the Probst bundle were the temporal association fibers, the splenium of the corpus

Table 3
Variability of fetal DTI measurements in different white matter structures across repeated scans. Variability is given as coefficient of variation of DTI metrics across the repeated scans, in % units. For each anatomical region, the following parameters are calculated: population mean ± standard deviation, value range and number of subjects (n) with suitable repeated measurement of the given ROI. Variability was calculated for only the subjects in whom the given structure was recognizable in at least two of the three repeated scans.

Name of region	Variability (%) of FA values	Variability (%) of MD values	Variability (%) of AD values	Variability (%) of RD values
Corpus callosum, genu	10.8 ± 7.7 (1.6–31.9) n = 19	6.1 ± 3.8 (0.2–13.3) n = 19	6.2 ± 3.29 (2.3–12.7) n = 19	6.7 ± 4 (0.29–13.9) n = 15
Corpus callosum, splenium	8.6 ± 6.1 (2.5–23.6) n = 16	6.9 ± 4.9 (0.2–17.4) n = 16	7 ± 4.9 (0.2–15.6) n = 16	7.2 ± 4.9 (0.15–18.6) n = 16
Internal capsule, posterior limb	8.2 ± 8.8 (2.1–40.4) n = 18	6.9 ± 6.1 (0.7–25.5) n = 18	6. 8 ± 6 (0.8–24.5) n = 18	7.2 ± 6.5 (0.5–26.1) n = 18
Brainstem fibers	14.6 ± 6.5 (2.9–23.8) n = 12	9.8 ± 10.4 (0.9–35.8) n = 12	10.9 ± 9.7 (1.7–31.5) n = 12	9.5 ± 10.8 (0.4–38.5) n = 12
Temporal and occipital association fibers	10.8 ± 9 (0.9–35.3) n = 18	6.6 ± 5.3 (0.7–23.4) n = 18	7.2 ± 5.8 (2.4–28) n = 18	6.5 ± 4.9 (0.6–20.5) n = 18
Probst's bundles	7.5 ± 3.9 (3.1–12.5) n = 4	4.4 ± 2.62 (0.6–6.7) n = 4	3 ± 2.1 (2.4–28) n = 4	5.5 ± 3.2 (1–8.6) n = 4

callosum and the internal capsule, posterior limb. While the genu and splenium of the corpus callosum were the two most frequently visible tracts across repeated scans, the variability of the DTI derived metrics was relatively high in these compared to the other structures.

The results for each white matter structure and each of the four investigated DTI derived metrics are summarized in Table 3, and for non-motion-corrected images in Table S2. We also investigated the distribution of variability across the population, which gives a better representation of the actual variability than the mean and standard deviations, due to the high skew of the distribution. Fig. 2 shows the frequency histograms of the variability of values and demonstrates that only a few fetuses had exceedingly high variability across repeated scans, (in some cases and structures > 20–25%), most likely due to acquisition related errors. With motion correction, the variability of values stayed under 10% for the majority of the subjects. The distribution of variability values also showed less scatter using motion correction and contained fewer extreme values, meaning that the motion correction procedure can ameliorate some of the effects leading to increased variability, at last in some cases.

3.3. Image variability, correlation and similarity

In Fig. 3 we summarized the results of the voxelwise analysis of value variability, voxel-to-voxel image correlations and structural similarity averaged for the entire brain. During the ROI based analysis, we concluded that the distribution of reproducibility is highly skewed due to outliers, *i.e.* fetuses whose DTI data are profoundly corrupted by artifacts in whom variability is very high. This observation was more pronounced for the whole-brain voxelwise analysis, most likely due to the fact that the DTI metrics calculated within the gray matter and the ventricular system show inherently higher variability. Therefore we decided to exclude 4 fetuses from this analysis, in which the mean framewise displacement was the highest (threshold at the 90th percentile: $FD_{mean} < 12$ mm).

After motion correction and exclusion of cases with the highest motion, the average image correlation across the repeated images was high, FA: R = 0.494 ± 0.23 (−0.0432–0.862), AD: R = 0.598 ± 0.258 (0.02–0.951), RD: R = 0.638 ± 0.269 (0.01–0.947) and MD: R = 0.632 ± 0.274 (0.076–0.949). Structural similarity was also high for all cases after motion correction, ranging from 0.909 to 0.998.

The motion correction procedure had a significant effect and resulted in lower standard deviation of FA values across the 3.0 T scans ($p = 0.0049$), higher image correlation of the FA maps (1.5 T cases: $p = 0.0203$, 3.0 T cases: $p = 0.0161$) and higher structural similarity of the FA images ($p = 0.0425$), tested using paired Wilcoxon signed rank tests. Controversially, the motion corrected RD and MD images had lower structural similarity index on 1.5 T ($p = 0.0017$ and $p = 0.0085$, respectively).

Scanner field strength only had a significant effect on the non-motion corrected axial diffusivity images on 1.5 T. The standard deviation of the AD maps was significantly higher on 3.0 T meaning lower reproducibility, $p = 0.0477$.

3.4. Factors influencing the reproducibility

Our study population included fetuses with normally appearing brain development and pathological cases. The presence of pathology affected the visibility of all fiber structures, they were less detectable in the pathological group, as indicated by the lower overall visibility score (normal brain development: 11.29 ± 5.56, pathologies: 7.83 ± 3.94). Fetuses with normally developing brains had lower variability of the FA of the temporooccipital association fibers and the MD of the splenium of the corpus callosum.

The effects of gestational age were pronounced the images that were not corrected for motion. Strong, positive correlations were evident between the gestational age of the fetus and the variability of FA, MD, RD and AD values of the brainstem and internal capsule. After fetal specific motion correction, the correlations between gestational age and variability were only moderate ($R^2 < 0.25$). Interestingly, larger fetuses tend to have larger reproducibility after the robust linear regression analysis, and the split between small (second trimester) and larger (third trimester) fetuses revealed that second trimester fetuses may have higher variability overall for the DTI derived measures.

Two MRI devices from the same vendor were used, which makes it possible to investigate the differences in reproducibility of DTI metrics on 1.5 T and 3.0 T field strengths, with nearly identical image acquisition parameters. The variability of the FA, MD and RD values of the temporooccipital association fibers were approximately twice as high on 3.0 as on 1.5 T, meaning lower reproducibility for these structures (Table 4).

The factor which seemed to influence the reproducibility measures the most was the mean displacement of the fetal head during the repeated scans. This parameter can be measured during the motion correction step and was expressed in mm scale (mean framewise displacement). Generally, moderate to strong correlations were found between fetal head motion and nearly all of the variability measures. In Table 4 we only report results if the adjusted R-squared value was higher than 0.25 (equals $R \geq 0.5$), which is considered strong correlation according to the rule of thumb by Cohen (Cohen, 1977), while the results of the entire analysis are detailed in Table S3. The greatest influence of motion on the variability was found before motion correction, and FA values were most affected by motion. Specifically, head motion was highly correlated with the variability of the FA within the internal capsule, posterior limb (correlation with FA variability: $R^2 = 0.68$). After motion correction, four repeatability values were still influenced by the original head motion: the MD, AD and RD values of the internal capsule, posterior limb ($R^2 = 0.612$, 0.605 and 0.531,

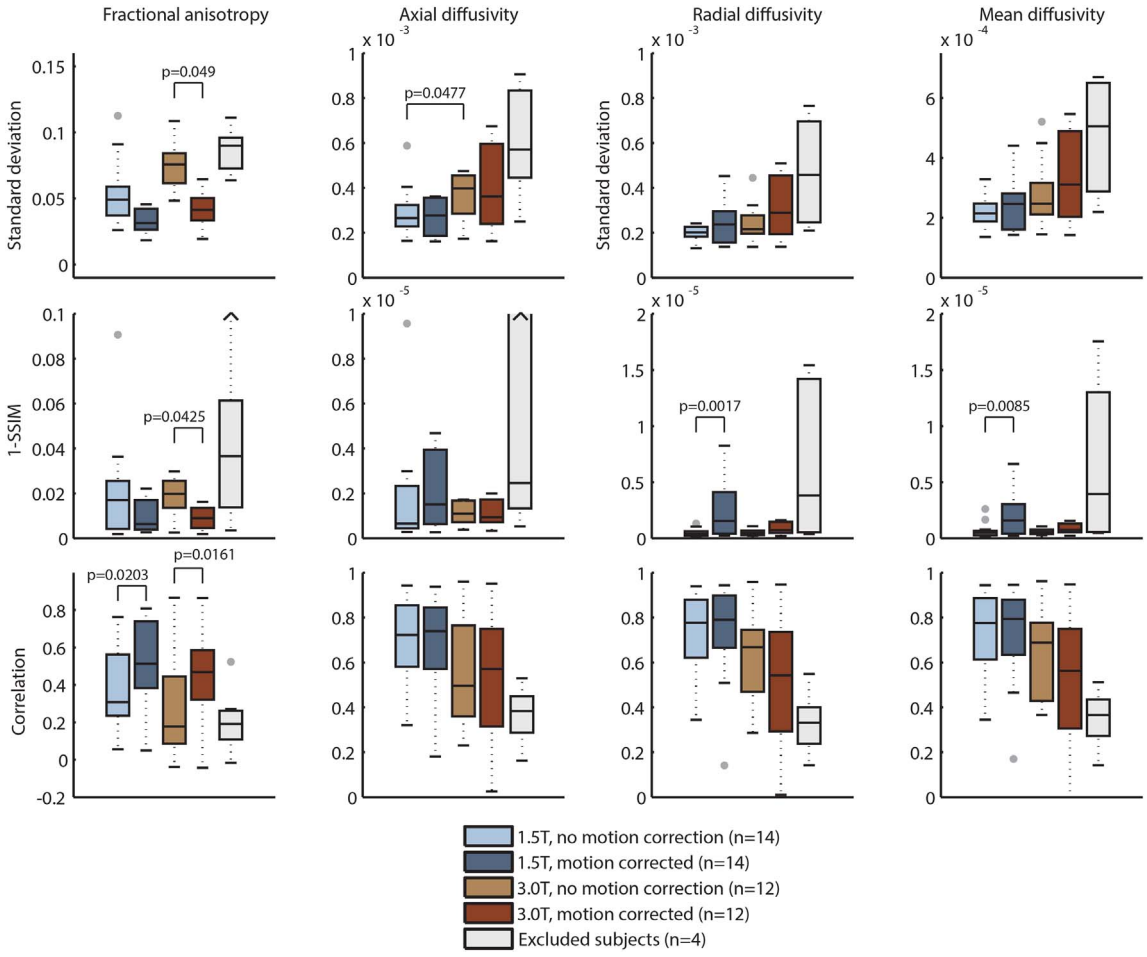

Fig. 3. Voxelwise reproducibility of DTI derived values based on whole-brain analysis. We calculated the standard deviation, image correlation and structural similarity index (depicted as 1-SSIM) across repeated scans. Each Whisker-plot demonstrates the distribution of the reproducibility metrics in cases that underwent fetal DTI on 1.5 T or 3.0 T, with and without fetal specific image processing including motion correction. The fifth box-plot demonstrates the variability of values in subjects that were excluded based on the criterion of excessive head motion during the DTI scan.

respectively). We summarized the influence of all investigated factors on the reproducibility measures in Table S4.

4. Discussion

We found that the visibility of major commissural, projection and association fibers in *in utero* DTI maps was moderate to high in 40–76% of the cases, but in a quarter of the cases at least one of the investigated structures was not visible in any of the three repeated scans, which is an important factor to consider for future studies investigating these structures. Unexpectedly, we did not find a close link between fetal age and the visibility of white matter bundles, which would be expected based on the known developmental trajectories of these structures. By the fetal age of 20 postmenstrual weeks, all the reported commissural and projection fibers are already formed (Dubois et al., 2013), however, none of the pathways have not yet reached full maturity, and most axons form synapses only at the level of the subplate (Kostović and Judas, 2010). Due to the restricted spatial resolution of the currently available *in utero* DTI, this level of detail of the axonal organization may remain undetectable. Furthermore, the presence of the radial migration pathways and the radial glia renders the subcortical zone highly anisotropic (Kolasinski et al., 2013), contributing to the limited visibility of some white matter structures in fetuses. Since long range association fibers have to “cross” the radial glia perpendicularly within the subcortical white matter, during mid-gestation this network of crossing fibers may restrict the ability of standard clinical DTI protocols to resolve the proper orientation of the principal diffusion vector, leading to a consistent lack of detectability in fetuses at a younger age. However, in the present study, the detectability of major white matter tracts was not significantly related to gestational age, so we assume that the differences in the detectability of major white matter structures arise as a result of technical limitations of the imaging method and do not necessarily reflect the underlying white matter developmental timing.

The Probst bundles were detected in all of the 5 CCA cases, and was reproducibly visible in 4 of these. These structures are identified as a large white matter structure running longitudinally in the horizontal (axial) plane, and which is easily distinguishable from the internal capsule, posterior limb at the level of the centrum semiovale based on the colored fractional anisotropy map (Fig. 2). The high visibility and high reproducibility of scalar values of this structure, even with such low case numbers, suggest that this bundle may be a promising prenatal maker of callosal agenesis, and *in utero* DTI may enable the characterization of how this ectopic bundle is formed (Kasprian et al., 2013; Mitter et al., 2015b; Jakab et al., 2015). Our findings show that the genu and splenium of the corpus callosum can also be detected with high accuracy on fetal DTI, however, the central part is usually not seen due to its thin cross-section. A viable strategy to increase the visibility of the central parts would be to decrease the slice thickness to 2 mm, which would enable the acquisition of isotropic pixels during DTI, at the cost of roughly doubling the imaging time. In order to confirm the presence of the central part of the corpus callosum, sagittal DTI acquisitions may help, optionally supported by super-resolution sampling techniques (Rousseau et al., 2010). By the same token, the visibility of the internal capsule, posterior limb may be improved by using coronal

Table 4
Factors significantly influencing the visibility of structures and the variability of *in utero* DTI measures across repeated scans. We evaluated whether gestational age, fetal head movement, presence of pathology or scanner field strength influences the visibility of tracts or the various variability measurements. R^2: coefficient of determination after linear regression as adjusted R-squared values, p: statistical significance of a Wilcoxon rank sum test. AD: axial diffusivity, FA: fractional anisotropy, GW: gestational weeks, MD: mean diffusivity, RD: radial diffusivity, SD: standard deviation.

Reproducibility parameter	Factor influencing reproducibility	Statistical values (adjusted-R^2, or p value)	Values across groups (mean ± SD, range)	
Variability of FA values, temporooccipital association fibers	Scanner field strength	p = 0.020	1.5 T 6.9% ± 5.0% (0.9%–14.9%)	3.0 T 17.1% ± 10.6% (5.4%–35.3%)
Variability of AD values, temporooccipital association fibers	Scanner field strength	p = 0.011	1.5 T 4.8% ± 1.9% (2.4%–8.2%)	3.0 T 10.9% ± 7.9% (4.5%–28.0%)
Variability of RD values, temporooccipital association fibers	Scanner field strength	p = 0.011	1.5 T 4.2% ± 2.7% (0.6%–7.4%)	3.0 T 10.1% ± 5.8% (5.5%–20.4%)
Overall visibility score of white matter structures	Presence of pathology	p = 0.039	Pathology present 7.83 ± 3.94 (1.00–13.00)	Unaffected brain development 11.29 ± 5.56 (1.00–15.00)
Variability of FA values, temporooccipital association fibers	Presence of pathology	p = 0.035	Pathology present 13.2% ± 9.5% (1.0%–35.3%)	Unaffected brain development 4.8% ± 2.9% (0.9%–8.1%)
Variability of MD values, corpus callosum, splenium	Presence of pathology	p = 0.042	Pathology present 8.9% ± 4.9% (0.0%–17.4%)	Unaffected brain development 3.6% ± 2.6% (0.4%–7.1%)
Visibility of corpus callosum, splenium	Presence of pathology	p = 0.023	Pathology present 0.41 ± 0.39 (0.00–1.00)	Unaffected brain development 0.81 ± 0.38 (0.00–1.00)
Variability of MD values, internal capsule, posterior limb	Head movement of fetus	R^2 = 0.612		
Variability of AD values, internal capsule, posterior limb	Head movement of fetus	R^2 = 0.605		
Variability of RD values, internal capsule, posterior limb	Head movement of fetus	R^2 = 0.531		
Visibility of corpus callosum, genu	Head movement of fetus	R^2 = 0.255		

acquisitions. The least visible structures in fetal DTI were the brainstem fibers, which were undetectable or not reproducible in 60% of the cases. This is most likely due to the confounding effect of the surrounding CSF spaces, the pulsation of tissues and the general fact that the cross-section of these fibers are small in the axial plane compared to the other investigated white matter structures. The lack of visibility of these structures underscores the need for care when judging the presence of brainstem or pons developmental abnormalities using DTI, as higher spatial and angular resolution is needed to more efficiently image these projection fibers.

In addition to the tract visibility, we also calculated the intrasubject, inter-session variability of the DTI metrics as an important indicator of the clinical usability of fetal DTI. According to our measurements, this aspect of reproducibility depends on the actual anatomical structure: while the corpus callosum and internal capsule had relatively low variability across scans, the brainstem fibers and the temporooccipital association bundles showed generally higher variability. This most likely stems from the fact that white matter with less mature fiber structure will have lower anisotropy and therefore lower signal to noise ratio during DTI scans. Furthermore, their size, their proximity to the ventricle system as a confounding factor may also influence the variability of the DTI metrics and the susceptibility of such structures to imaging artifacts.

Before the DTI metrics can be considered as viable prenatal markers of white matter integrity and development, it is important to evaluate their reproducibility based on previously reported data. The reproducibility of DTI derived metrics has been estimated previously in studies of adult volunteers, but such published data cannot be directly compared to our results due to the different study population and the fact that typical fetal DTI protocols differ in design to the optimal adult protocols, due to the requirement for short scan times. Such protocol optimization for fetal DTI usually involves reducing the number of diffusion-weighted gradient directions and the number of image slices. Although six diffusion-weighted magnetic field gradient directions may be sufficient for estimating the FA and MD values (Lebel et al., 2012), there is a clear advantage associated with using 15 or 30 directions (Wang et al., 2012). DTI in adults with at least 30 diffusion-weighting directions according to the Jones-30 scheme was reported to be optimal for reproducibly estimating the fractional anisotropy and mean diffusivity (Jones et al., 1999), however, similar investigations for fetuses have not yet been done. The reproducibility of DTI measurements is inferior to that reported in adult volunteers; however, the coefficient of variation (CV) for more than half of the subjects in our study lies in a similar range to that reported in the following adult studies.

Using the 30 direction acquisition scheme, the CV of the voxel-based and ROI-based diffusion metrics were below 10% for a similar set of white matter structures to those investigated in our study (Farrell et al., 2007), while gray matter structures, *i.e.* areas of low fractional anisotropy, showed considerably higher variability. Within scanner variability of the FA and MD values of the corpus callosum were reported to be as low as 1.9% and 2.6% (Pfefferbaum et al., 2003), while Cercignani and colleagues found that histogram-derived metrics of mean diffusivity were highly reproducible (coefficients of variation ranging from 1.72% to 5.56%), as were fractional anisotropy histogram-derived metrics (coefficients of variation ranging from 5.45% to 7.34%) (Cercignani et al., 2003). A multicenter study reported CVs of 2.2% for ADC, 3.5% for axial diffusivity and 8.7% for FA in the global white matter over time on the same scanner (Fox et al., 2012), while in children, Bonekamp et al. reported a between-scan reproducibility ranging from 2.6% to 4.6% for the FA and from 0.8% to 3.4% for the ADC (Bonekamp et al., 2007). Both observations are consistent with our finding of inferior reproducibility of the direction-dependent metrics, such as AD compared to rotationally invariant measures like the MD. Using 12 diffusion weighting directions, Heiervang and colleagues have shown that inter-session and inter-subject CVs for the FA in the cingulum bundle, pyramidal tracts, optic radiations and genu of the corpus callosum were 5% and 10%, respectively, while inter-session and inter-subject CVs for the MD were below 3% and 8%, respectively (Heiervang et al., 2006). In a report from Wang et al., out of 60 tractography measurements, 43 showed intersession CV ≤ 10%, and the most reliable regions were the corpus callosum, cingulum, cerebral peduncles, the uncinate and the arcuate fascicle (Wang et al., 2012).

The effect of differences in scanner design and MRI vendor on the variability of fetal DTI is also an important factor to consider in addition to the intravendor (intrasubject) variability, as the variability in DTI metrics depends considerably on the coil systems, imagers, vendors, and field strengths used for MRI (Sasaki et al., 2008). In a study by Pfefferbaum et al., DTI measures were reported to show a systematic bias across scanners (CV = 4.5% for FA and CV = 7.5% for trace) (Pfefferbaum et al., 2003). A study across 5 scanners also reported higher inter-scanner variability than within-scanner variability across time (Fox et al., 2012), and in another similar study, the intra-site CV for FA ranged from 0.8% to 3.0%, while the inter-site CV ranged from 1.0% to 4.1% (Vollmar et al., 2010). Validation experiments suggest that DTI protocols with sufficiently low coefficient of variation not only allow for the estimation of per-voxel estimates of diffusion, but also the reconstruction of fiber pathways reproducibly using tractography techniques, usually within one study site (Heiervang et al., 2006; Wakana et al., 2007; Malykhin et al., 2008; Danielian et al., 2010; Besseling et al., 2012; Wang et al., 2012). However, we note that the higher inter-scanner variability (relative to the intra-subject variability) becomes more pronounced during diffusion tractography and causes less comparable results across sites (Jakab et al., 2016), and this inter-scanner variability should also be considered when performing tractography using fetal DTI data.

In the present study, we observed low standard deviations of DTI metrics across the scans, high image correlation and high structural similarity with voxel based whole-brain analysis of the images, and a similar trend was apparent in the ROI-based reproducibility and visibility analysis. While motion correction and the exclusion of severely corrupted data sets improves data reproducibility (Fig. 3), the 3.0 T DTI acquisitions appear to be inferior in comparison to the 1.5 T data. We observed better reproducibility results when using a voxel-level comparison of the repeated images. The lower variability compared to the ROI approach, high image similarity and correlation for all image voxels, including gray matter and cerebrospinal fluid pixels, may not necessarily reflect improved data usability. Artifacts that have an influence for all brain voxels or large parts of the brain equally can result in inflated per-pixel correlations between scans, while structural similarity may not accurately reflect the absolute variability of values across repeated scans. From previous studies reported in the literature, an inter-session DTI metric variability of < 10% appears to be a ubiquitous finding in adults or in children, which is only comparable to the *in utero* measurements if all fetal scans demonstrating severe artifacts are consistently excluded, leading to a data loss affecting at least 15% to even 25% of the subjects.

Our study allows us to estimate the effect of various factors on the reproducibility of *in utero* DTI, including scanner field strength, the presence of pathology and fetal head movements. Significant differences in white matter structure visibility between 1.5 and 3.0 T were found even when fetal-specific image processing was applied. In this case, the variability of diffusion values of the temporooccipital association tracts was higher on the 3.0 T data. As 3.0 T is being increasingly used for fetal imaging (Welsh et al., 2011) due to the higher spatial resolution and better tissue contrast in T2-weighted imaging associated with the higher field strength, this finding raises concerns. A possible explanation for the lower tract reproducibility on 3.0 T fetal DTI is that despite the increased SNR and reports of higher reproducibility with higher field strengths in adults (Polders et al., 2011), susceptibility related artifacts arising from the heterogeneous environment surrounding the fetus lead to more pronounced field inhomogeneities, as described in previous reports (Rosenkrantz et al., 2011; Lavdas et al., 2014). To reduce the inter-slice effects arising in EPI acquisitions and the influence of field inhomogeneity, various approaches have been suggested (Keraudren et al., 2014; Fogtmann et al., 2014; Ferrazzi et al., 2014), of which we adapted a normalization of the Z-profile of the intensities on a slice-wise basis, in order to improve our estimate of the diffusion anisotropy. Our study sample is broadly representative of the wider clinical fetal MRI population, in which 0–25% of the cases are typically evaluated as showing normal brain development. This allowed us to judge if DTI metrics from normally developing fetuses have different reproducibility from those estimated in fetuses with intracerebral pathology. The presence of pathology was associated with a worse visibility of fibers overall and the visibility of the splenium of the corpus callosum, which is most likely due to the fact that pathological fetuses often had hydrocephalus, which causes thinning (due to physical compression) of the periventricular fiber structures, resulting in lower visibility when imaged with a large voxel size and slice thickness.

Not surprisingly, the degree of fetal head movement during the scan, the mean framewise displacement, was positively correlated with the variability of nearly all investigated structures (Tables 4 and S3). Real motion – *i.e.*, the displacement of the head relative to its environment – can stem from fetal head and trunk movements and maternal breathing, while apparent motion of the brain can also arise from susceptibility related distortions, which are more pronounced *in utero*. These distortions cause spin history artifacts that influence the signal intensity (Ferrazzi et al., 2014). Pulsation from the amniotic fluid and surrounding organs, such as the maternal aorta, may further lead to spin dephasing during the diffusion-weighting sequence. The displacement of voxels between the individual diffusion-weighted image frames corrupts the reconstruction of the diffusion tensors and results in lower FA values and incorrect orientations of the calculated eigenvectors (Aksoy et al., 2010), consistent with our observation that the mean framewise displacement mostly affected the directionally dependent axial and radial diffusivity.

Our study further enables us to outline two different strategies to tackle the problem of lower reproducibility of current, clinically viable *in utero* DTI sequences. First, the retrospective correction of fetal head movement can be applied to achieve more reproducible reconstruction of the diffusion tensor, for which we used a custom image processing work-flow in our study (Supplementary document). Unintentional head motion is a common problem in imaging studies, and various reconstruction approaches have been suggested for fetal functional MRI (Fogtmann et al., 2014), as well as for fetal and neonatal DTI (Jiang et al., 2009). A common way to improve image quality is to account for fetal head motion by re-aligning each time frame to a selected reference point (Malamateniou et al., 2013). This realignment step is implemented in common functional MRI processing tools, however, its use in fetal diffusion imaging is not yet established. Retrospective correction of motion was beneficial for our study, which is consistent with findings from the literature (Holdsworth et al., 2012). A particularly promising approach is to reconstruct fetal diffusion-weighted MRI data on regular grids from scattered data (Oubel et al., 2012). However, it also appears that fetal head motion reduces reproducibility slightly even if the image frames and diffusion vectors are corrected for motion, and hence further mitigation strategies are needed. Another solution to improve data usability is to collect the DTI in multiple acquisitions spread across the MRI examination. According to our experience, excessive fetal movements that lead to considerable signal loss during DTI may periodically increase or decrease in activity multiple times during a typical, hour-long fetal MRI. It is not possible to forecast such periods, but we can increase the chance for a motion-free DTI by leaving time between the scans. To maximize the clinical output during the “waiting periods”, other clinical sequences designed for observing fetal movements (Brugger et al., 2006), or fast T2-weighted anatomical scans may be acquired. As the fetus may rotate its head significantly in-between the waiting periods, re-orientation of the b-matrix is required during the fetal-specific image post processing (Supplementary document). Repeated DTI has further benefits. Multiple, interleaved B_0 images allow for the estimation of the noise floor for each pixel, enabling a more robust estimation of the diffusion tensor, such as that using the RESTORE algorithm (Chang et al., 2005). The rotated b-matrix means that images are acquired with slightly different diffusion-weighting orientations, which allows for a subsampling of more non-collinear

directions to achieve higher angular resolution. Therefore, by utilizing within-subject repeated DTI acquisitions with clinically feasible imaging parameters within a single scanning session, the fiber architecture can be reconstructed in finer, three-dimensional detail, following the "super-resolution" approach (Rousseau et al., 2010; Fogtmann et al., 2014; Rousseau et al., 2006).

Our prenatal neuroimaging study suffers from a number of important limitations. The statistical power is limited by the exploratory nature of the study, and specifically the small participant group sizes. Increasing the case numbers based on clinical data, especially for fetuses with normal brain development, is challenging due to the time constraints of fetal MRI and the relatively low number of such scans. In addition, the participant group is heterogeneous in terms of gestational age, and the brain anatomy is often affected by various pathologies, such as hydrocephalus, asymmetric lateral ventricle size and associated developmental abnormalities. A further limiting factor originates from the fact that fetal MRI protocols often have to make a compromise between resolution, slice thickness, fast imaging time, patient comfort, reduced SAR, and artifacts arising from fetal motion. Our study is based on a relatively thick-slice DTI protocol, with 15 diffusion weighting directions and a b-factor of 700 s/mm^2. In particular, the large slice thickness in fetal studies is likely to present a confounding factor when testing reproducibility, since smaller white matter bundles, especially if they run parallel to the imaging plane, may be only partially imaged and good anatomical correspondence between participants is hard to achieve. This effect is more pronounced if the fetus moved or rotated its head perpendicular to the axial plane in any of the image frames, but may be partially mitigated by oversampling the DTI from different directions, as proposed in super-resolution studies for structural and diffusion imaging (Rousseau et al., 2010; Oubel et al., 2012).

A considerable physiological limitation of the current study is that the exact microstructural correlates of the described white matter bundles are unknown. It is known that in DTI, the majority of the anisotropy stems from the microscopic structure of the axonal membranes (Beaulieu, 2002), and to a lesser extent, from the myelin sheath around the axons. Elements of the extracellular matrix, axonal tubules or other, even non-neural structures may only have a limited role in causing anisotropic diffusion. During the fetal age range that was used in our study, however, neither neuronal migration, axonal growth, pathfinding, nor myelination are complete. We cannot rule out the possibility that transient cellular components, such as radial glia may modify the DTI measurements *in utero*. The clarification of such confounds would require future work with histological work-up in fetal specimens, including further works comparing DTI results *in utero* and *post mortem* and analyzing fiber structures in histological samples (Mitter et al., 2015b).

We conclude that the reproducibility of *in utero* DTI is comparable to that of adult studies only if we exclude the subjects whose images were severely compromised by artifacts and if we utilize fetal image specific post processing approaches. This observation provides a note of caution for studies attempting to use *in utero* DTI as a marker of disease in individual clinical cases: even with scans repeated three times, some white matter fiber bundles can remain undetectable and the entire session may be corrupted from artifacts that are not mitigated with image post processing. However, in group studies, *in utero* DTI may offer a promising approach for depicting white matter anatomy and measuring microstructural properties of tissue diffusion, and could even allow for the reconstruction of the emerging brain connectivity prenatally. Future directions of fetal imaging research should emphasize the more robust reconstruction of the diffusion tensor based on repeated data and higher angular resolution acquisition schemes. In this endeavor, fetal-specific image processing and repeated scanning is recommended to ensure the detectability of white matter structures.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.nicl.2017.06.013.

Acknowledgements

A. J. was supported by the European Union FP7 Marie Curie IEF Research grant FABRIC—"Exploring the Formation and Adaptation of the Brain Connectome," grant no. 2012-PIEF-GA-33003. A.J. and R.O.T. are supported by the EMDO Foundation, grant number 928.

References

Aksoy, M., Skare, S., Holdsworth, S., et al., 2010. Effects of motion and b-matrix correction for high resolution DTI with short-axis PROPELLER-EPI. NMR Biomed. 23 (7), 794–802.
Arfanakis, K., Haughton, V.M., Carew, J.D., et al., 2002. Diffusion tensor MR imaging in diffuse axonal injury. AJNR Am. J. Neuroradiol. 23 (5), 794–802.
Babikian, T., Tong, K.A., Galloway, N.R., et al., 2009. Diffusion-weighted imaging predicts cognition in pediatric brain injury. Pediatr. Neurol. 41 (6), 406–412.
Bar-Shir, A., Duncan, I.D., Cohen, Y., 2009. QSI and DTI of excised brains of the myelin-deficient rat. NeuroImage 48 (1), 109–116.
Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. B 111 (3), 209–219.
Basser, P.J., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B. 103 (3), 247–254.
Basser, P.J., Pajevic, S., Pierpaoli, C., et al., 2000. In vivo fiber tractography using DT-MRI data. Magn. Reson. Med. 44 (4), 625–632.
Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system? A technical review. NMR Biomed. 15 (7–8), 435–455.
Besseling, R.M., Jansen, J.F., Overvliet, G.M., et al., 2012. Tract specific reproducibility of tractography based morphology and diffusion metrics. PLoS One 7 (4), e34125.
Bonekamp, D., Nagae, L.M., Degaonkar, M., et al., 2007. Diffusion tensor imaging in children and adolescents: reproducibility, hemispheric, and age-related differences. NeuroImage 34 (2), 733–742.
Brugger, P.C., Mittermayer, C., Prayer, D., 2006. A new look at the fetus: thick-slab T2-weighted sequences in fetal MRI. Eur. J. Radiol. 57 (2), 182–186.
Catani, M., Howard, R.J., Pajevic, S., et al., 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. NeuroImage 17 (1), 77–94.
Cercignani, M., Bammer, R., Sormani, M.P., et al., 2003. Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. AJNR Am. J. Neuroradiol. 24 (4), 638–643.
Chang, L.C., Jones, D.K., Pierpaoli, C., 2005. RESTORE: robust estimation of tensors by outlier rejection. Magn. Reson. Med. 53 (5), 1088–1095.
Cohen, J., 1977. Chapter 1 - the concepts of power analysis. In: Cohen, J. (Ed.), Statistical Power Analysis for the Behavioral Sciences (Revised Edition). Academic Press, pp. 1–17. http://dx.doi.org/10.1016/B978-0-12-179060-8.50006-2.
Danielian, L.E., Iwata, N.K., Thomasson, D.M., et al., 2010. Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. NeuroImage 49 (2), 1572–1580.
Doria, V., Beckmann, C.F., Arichi, T., Merchant, N., Groppo, M., Turkheimer, F.E., Counsell, S.J., Murgasova, M., Aljabar, P., Nunes, R.G., Larkman, D.J., Rees, G., Edwards, A.D., 2010. Emergence of resting state networks in the preterm human brain. Proc. Natl. Acad. Sci. U. S. A. 107 (46), 20015–20020 (Nov 16).
Dubois, J., Dehaene-Lambertz, G., Kulikova, S., et al., 2013. The early development of brain white matter: a review of imaging studies in fetuses, newborns and infants. Neuroscience 276, 48–71.
Farrell, J.A., Landman, B.A., Jones, C.K., et al., 2007. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. J. Magn. Reson. Imaging 26 (3), 756–767.
Ferrazzi, G., Kuklisova Murgasova, M., Arichi, T., et al., 2014. Resting State fMRI in the moving fetus: a robust framework for motion, bias field and spin history correction. NeuroImage 101, 555–568.
Fogtmann, M., Seshamani, S., Kroenke, C., et al., 2014. A unified approach to diffusion direction sensitive slice registration and 3-D DTI reconstruction from moving fetal brain anatomy. IEEE Trans. Med. Imaging 33 (2), 272–289.
Fox, R.J., Sakaie, K., Lee, J.C., et al., 2012. A validation study of multicenter diffusion tensor imaging: reliability of fractional anisotropy and diffusivity values. AJNR Am. J. Neuroradiol. 33 (4), 695–700.
Heiervang, E., Behrens, T.E., Mackay, C.E., et al., 2006. Between session reproducibility and between subject variability of diffusion MR and tractography measures. NeuroImage 33 (3), 867–877.
Hess, C.P., 2009. Update on diffusion tensor imaging in Alzheimer's disease. Magn. Reson. Imaging Clin. N. Am. 17 (2), 215–224.
Holdsworth, S.J., Aksoy, M., Newbould, R.D., et al., 2012. Diffusion tensor imaging (DTI) with retrospective motion correction for large-scale pediatric imaging. J. Magn. Reson. Imaging 36 (4), 961–971.
Huang, H., Xue, R., Zhang, J., et al., 2009. Anatomical characterization of human fetal brain development with diffusion tensor magnetic resonance imaging. J. Neurosci. 29 (13), 4263–4273.
Jakab, A., Kasprian, G., Schwartz, E., et al., 2015. Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. NeuroImage 111, 277–288.
Jakab, A., Werner, B., Piccirelli, M., et al., 2016. Feasibility of diffusion tractography for the reconstruction of intra-thalamic and cerebello-thalamic targets for functional neurosurgery: a multi-vendor pilot study in four subjects. Front. Neuroanat. 10, 76.

Jiang, S., Xue, H., Counsell, S., et al., 2009. Diffusion tensor imaging (DTI) of the brain in moving subjects: application to in-utero fetal and ex-utero studies. Magn. Reson. Med. 62 (3), 645–655.
Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn. Reson. Med. 42 (3), 515–525.
Kasprian, G., Brugger, P.C., Weber, M., et al., 2008. In utero tractography of fetal white matter development. NeuroImage 43 (2), 213–224.
Kasprian, G., Del Rio, M., Prayer, D., 2010. Fetal diffusion imaging: pearls and solutions. Top. Magn. Reson. Imaging 21 (6), 387–394.
Kasprian, G., Brugger, P.C., Schopf, V., et al., 2013. Assessing prenatal white matter connectivity in commissural agenesis. Brain 136 (Pt 1), 168–179.
Keraudren, K., Kuklisova-Murgasova, M., Kyriakopoulou, V., et al., 2014. Automated fetal brain segmentation from 2D MRI slices for motion correction. NeuroImage 101, 633–643.
Kolasinski, J., Takahashi, E., Stevens, A.A., et al., 2013. Radial and tangential neuronal migration pathways in the human fetal brain: anatomically distinct patterns of diffusion MRI coherence. NeuroImage 79, 412–422.
Kostović, I., Judas, M., 2010. The development of the subplate and thalamocortical connections in the human foetal brain. Acta Paediatr. 99 (8), 1119–1127. http://dx.doi.org/10.1111/j.1651-2227.2010.01811.x. (Aug, Epub 2010 Mar 29. Review. PubMed PMID: 20367617).
Landman, B.A., Farrell, J.A., Jones, C.K., et al., 2007. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. NeuroImage 36 (4), 1123–1138.
Lavdas, I., Miquel, M.E., McRobbie, D.W., et al., 2014. Comparison between diffusion-weighted MRI (DW-MRI) at 1.5 and 3 tesla: a phantom study. J. Magn. Reson. Imaging 40 (3), 682–690.
Le Bihan, D., Johansen-Berg, H., 2012. Diffusion MRI at 25: exploring brain tissue structure and function. NeuroImage 61 (2), 324–341.
Le Bihan, D., Breton, E., Lallemand, D., et al., 1988. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 168 (2), 497–505.
Lebel, C., Benner, T., Beaulieu, C., 2012. Six is enough? Comparison of diffusion parameters measured using six or more diffusion-encoding gradient directions with deterministic tractography. Magn. Reson. Med. 68 (2), 474–483.
Lu, S., Ahn, D., Johnson, G., et al., 2003. Peritumoral diffusion tensor imaging of high-grade gliomas and metastatic brain tumors. AJNR Am. J. Neuroradiol. 24 (5), 937–941.
Malamateniou, C., Malik, S.J., Counsell, S.J., et al., 2013. Motion-compensation techniques in neonatal and fetal MR imaging. AJNR Am. J. Neuroradiol. 34 (6), 1124–1136.
Malykhin, N., Concha, L., Seres, P., et al., 2008. Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. Psychiatry Res. 164 (2), 132–142.
Mitter, C., Prayer, D., Brugger, P.C., et al., 2015a. In vivo tractography of fetal association fibers. PLoS One 10 (3), e0119536.
Mitter, C., Jakab, A., Brugger, P.C., et al., 2015b. Validation of in utero tractography of human fetal commissural and internal capsule fibers with histological structure tensor analysis. Front. Neuroanat. 9, 164.
Oubel, E., Koob, M., Studholme, C., et al., 2012. Reconstruction of scattered data in fetal diffusion MRI. Med. Image Anal. 16 (1), 28–37.
Owen, J.P., Ziv, E., Bukshpun, P., et al., 2013. Test-retest reliability of computational network measurements derived from the structural connectome of the human brain. Brain Connect. 3 (2), 160–176.
Pfefferbaum, A., Adalsteinsson, E., Sullivan, E.V., 2003. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. J. Magn. Reson. Imaging 18 (4), 427–433.
Polders, D.L., Leemans, A., Hendrikse, J., et al., 2011. Signal to noise ratio and uncertainty in diffusion tensor imaging at 1.5, 3.0, and 7.0 Tesla. J. Magn. Reson. Imaging 33 (6), 1456–1463.
Price, S.J., 2007. The role of advanced MR imaging in understanding brain tumour pathology. Br. J. Neurosurg. 21 (6), 562–575.
Raab, P., Hattingen, E., Franz, K., et al., 2010. Cerebral gliomas: diffusional kurtosis imaging analysis of microstructural differences. Radiology 254 (3), 876–881.
Rosenkrantz, A.B., Oei, M., Babb, J.S., et al., 2011. Diffusion-weighted imaging of the abdomen at 3.0 Tesla: image quality and apparent diffusion coefficient reproducibility compared with 1.5 Tesla. J. Magn. Reson. Imaging 33 (1), 128–135.
Rousseau, F., Glenn, O.A., Iordanova, B., et al., 2006. Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. Acad. Radiol. 13 (9), 1072–1081.
Rousseau, F., Kim, K., Studholme, C., et al., 2010. On super-resolution for fetal brain MRI. Med. Image Comput. Comput. Assist Interv. 13 (Pt 2), 355–362.
Sasaki, M., Yamada, K., Watanabe, Y., et al., 2008. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology 249 (2), 624–630.
Sotiropoulos, S.N., Jbabdi, S., Xu, J., et al., 2013. Advances in diffusion MRI acquisition and processing in the human connectome project. NeuroImage 80, 125–143.
Sporns, O., 2011. The human connectome: a complex network. Ann. N. Y. Acad. Sci. 1224 (1), 109–125.
Tropine, A., Vucurevic, G., Delani, P., et al., 2004. Contribution of diffusion tensor imaging to delineation of gliomas and glioblastomas. J. Magn. Reson. Imaging 20 (6), 905–912.
Vollmar, C., O'Muircheartaigh, J., Barker, G.J., et al., 2010. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. NeuroImage 51 (4), 1384–1394.
Wakana, S., Caprihan, A., Panzenboeck, M.M., et al., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage 36 (3), 630–644.
Wang, Z., Bovik, A.C., Sheikh, H.R., et al., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.
Wang, J.Y., Abdi, H., Bakhadirov, K., et al., 2012. A comprehensive reliability assessment of quantitative diffusion tensor tractography. NeuroImage 60 (2), 1127–1138.
Welsh, R.C., Nemec, U., Thomason, M.E., 2011. Fetal magnetic resonance imaging at 3.0T. Top. Magn. Reson. Imaging 22 (3), 119–131.
Westin, C., Maier, S.E., Mamata, H., et al., 2002. Processing and visualization for diffusion tensor MRI. Med. Image Anal. 6 (2), 93–108.