A multi-center study: Intra-scan and inter-scan variability of diffusion spectrum imaging

A. Lemkaddem a,⁎, A. Daducci a, S. Vulliemoz b,g, K. O'Brien d,i, F. Lazeyras d,i, M. Hauff f, R. Wiest f, R. Meuli c,h, M. Seeck b,g, G. Krueger e, J.-P. Thiran a,c,h

a Ecole Polytechnique Fédéral de Lausanne, Signal Processing Laboratories (LTS5), Lausanne, Switzerland
b Epilepsy Unit, Neurology Clinic, University Hospitals of Geneva, Switzerland
c University Hospital Center (CHUV), Lausanne, Switzerland
d Service of Radiology, University of Geneva, Geneva, Switzerland
e Siemens Schweiz AG, Healthcare Sector IM&WS S, Renens, Switzerland
f Support Center for Advanced Neuroimaging, Institute of Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, University of Bern, Switzerland
g Faculty of Medicine of Geneva, Switzerland
h University of Lausanne (UNIL), Switzerland
i University Hospitals of Geneva, Geneva, Switzerland

ABSTRACT

The objective of this study was to investigate whether it is possible to pool together diffusion spectrum imaging data from four different scanners, located at three different sites. Two of the scanners had identical configuration whereas two did not. To measure the variability, we extracted three scalar maps (ADC, FA and GFA) from the DSI and utilized a region and a tract-based analysis. Additionally, a phantom study was performed to rule out some potential factors arising from the scanner performance in case some systematic bias occurred in the subject study. This work was split into three experiments: intra-scanner reproducibility, reproducibility with twin-scanner settings and reproducibility with other configurations. Overall for the intra-scanner and twin-scanner experiments, the region-based analysis coefficient of variation (CV) was in a range of 1%–4.2% and below 3% for almost every bundle for the tract-based analysis. The uncinate fasciculus showed the worst reproducibility, especially for FA and GFA values (CV 3.7–6%). For the GFA and FA maps, an ICC value of 0.7 and above is observed in almost all the regions/tracts. Looking at the last experiment, it was found that there is a very high similarity of the outcomes from the two scanners with identical setting. However, this was not the case for the two other imagers. Given the fact that the overall variation in our study is low for the imagers with identical settings, our findings support the feasibility of cross-site pooling of DSI data from identical scanners.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Multi-center neuroimaging studies may provide a powerful framework to conduct comprehensive studies of basic neuroanatomy and clinical disorders for research, clinical and pharmaceutical purposes. Since patient access can be significantly increased, multi-center studies allow a substantial speed up of research and development projects. Multi-center studies also provide the opportunity to introduce standards that could move MRI toward a quantitative and user-independent diagnostic tool.

Collecting and combining images from different imaging devices, however, introduce additional sensitivity to systematic variability in the data. Variability may arise from device-dependent factors, differences in the reconstruction and processing strategies, in the case of in-vivo measurements, and also from physiological aspects that to the end might confound the interpretation of the results.

Since the introduction of diffusion magnetic resonance imaging (dMRI) (Le Bihan and Breton, 1985), the method has evolved to a standard neuroimaging protocol in clinical diagnosis. During the past decade, dMRI has also been increasingly exploited in clinical and research studies as it allows assessing the micro-structural integrity of the brain.

Some dMRI-derived parameters have been proven to be a sensitive marker of disease-related structural changes for many pathologies and disorders. For example, a reduction of the apparent diffusion coefficient (ADC) has been demonstrated to be a clinical marker of tissue death (Roberts and Rowley, 2003), while changes in the fractional anisotropy (FA) were used as a biomarker in diseases like epilepsy (Focke et al., 2008; Schoene-Bake et al., 2009), multiple sclerosis (Ge et al., 2005; Goldberg-Zimring et al., 2005), schizophrenia (Sun et al., 2003), and even in normal aging
(Pfefferbaum et al., 2000) and the developing brain (Bonekamp et al., 2007).

The most common analysis approach used in dMRI group studies consists in defining regions of interest across the brain and then comparing the groups based on the values of some dMRI-derived measures, such as FA or ADC, in those regions (Bonekamp et al., 2007; Lau and Goodyear, 2007; Vollmar et al., 2010). An alternative connectivity-oriented approach is to use tractography (Basser et al., 2000) to compute fiber pathways through the white matter and then considering the values of the scalar measures (i.e. ADC, FA etc.) along those pathways (Concha et al., 2010; Kanaan et al., 2006; Pannek et al., 2011).

As these techniques are potentially of high clinical value and increasingly used in clinical research, it is very important to characterize and understand their reproducibility in detail. So far, however, reproducibility in diffusion MRI has obtained only little attention and has been studied only in diffusion tensor imaging (DTI) (Ciccarelli et al., 2003; Heiervang et al., 2006; Marencio et al., 2006). Unfortunately, there are only very few studies addressing the reliability of cross-center measures at 3 T (Fox et al., 2011; Teipel et al., 2011; Vollmar et al., 2010; Zhu et al., 2011), that would allow us to draw generalized conclusion about the dMRI reproducibility and in particular to understand the most limiting factors in cross-scanner comparisons.

Moreover, high angular resolution diffusion imaging (HARDI) schemes and in particular diffusion spectrum imaging (DSI) are becoming increasingly available and attractive as these schemes are capable of measuring more accurate white matter fiber tracts and are increasingly used to study structural connectivity in the brain (Gigandet et al., 2008; Granziera et al., 2009, 2011; Hagmann et al., 2008; Honey et al., 2009; Knock et al., 2009; Wedeen et al., 2005). To our knowledge, no reports of multi-site reproducibility from more advanced dMRI acquisition schemes have been published.

In the context of a Swiss multi-center study involving 3 different hospitals and 4 scanners and aiming at measuring structural connectivity alterations in a cohort of epileptic patients, we acquired DSI MRI data to explore the following questions: (i) what are the factors determining intra- and inter-scanner variability in DSI data, (ii) how can we assess the intra- and inter-scanner variability in DSI measurements, (iii) are intra- and inter-scanner variations low enough to permit pooling of clinical dMRI data and (iv) what is the relation between diffusion tensor related FA and HARDI related GFA (generalized fractional anisotropy) measures.

Materials and methods

MRI data was collected from four 3 T MRI scanners from the same vendor (Siemens, Erlangen, Germany) located at three different imaging centers. Table 1 summarizes the differences in technical specifications. Scanners A and B (both Magnetom Trio a Tim System) are equipped with a 32-channel head helmet coil and will be referred to as “twin scanners” in the following. Scanner C (Magnetom Verio) is also equipped with a 32-channel head helmet coil but is based on a 70 cm magnet bore opening. Scanner D (Magnetom Trio, a Tim System) is equipped with a 12-channel head matrix coil. Scanners C and D are located in the same hospital.

Table 1 Details of scanner configurations.

<table>
<thead>
<tr>
<th>Scanner</th>
<th>Model</th>
<th>RF-coil</th>
<th>Coil combine</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Trio</td>
<td>32</td>
<td>Adaptive-combine</td>
</tr>
<tr>
<td>B</td>
<td>Trio</td>
<td>32</td>
<td>Adaptive-combine</td>
</tr>
<tr>
<td>C</td>
<td>Verio</td>
<td>32</td>
<td>Sum-of-square</td>
</tr>
<tr>
<td>D</td>
<td>Trio</td>
<td>12</td>
<td>Adaptive-combine</td>
</tr>
</tbody>
</table>

Recent studies have pointed out the dependency of FA and ADC in the brain tissue on the underlying signal-to-noise-ratio (SNR) (Burdette et al., 2001). Thus, SNR differences among scanners may introduce inter-scanner variations and we consider a basic SNR measurement a meaningful criterion for qualification. Furthermore differences in eddy-current and shim performance between scanners may introduce inter-scanner and inter-session variations. Thus, a further qualification criterion could be the assessment of the eddy-current performance and the obtained B0-homogeneity. It should be noted, that those factors can be corrected if appropriate information and algorithms for artifact correction are applied.

Phantom study

SNR-measurements

SNR and eddy current (EC) performance of the scanners were assessed using a Siemens spherical imaging phantom (1000 g of distilled water doped with 1.25 g NiSO4.6H2O) that is also used up for the service tune-up procedure. Since we used in our study multi-coil receive head arrays and parallel imaging with an acceleration of 2, the obtained SNR strongly depends on the position of the object in the coil coordinate system. To obtain a pixel-wise SNR measurement, we acquired a time-series of images without diffusion weighting (b-value=0 s/mm²) using a single-shot echo-planar imaging (EPI) product sequence consisting of 22 interleaved slices with 32 repetitions. The other image parameters of the sequence were: TR/TE=4200/154 ms, voxel size = 2.2×2.2×3 mm, and matrix size = 96×96. The SNR of this EPI time-series, E (x, y, z, t), where t = 1, ..., 32, was computed for each voxel as follows (Glover and Lai, 1998):

\[
SNR(x,y,z) = \frac{\mu(S)}{\sigma(N)}
\]

where the signal and the noise are estimated as

\[
S(x,y,z,1:16) = \frac{[E(x,y,z,2i) + E(x,y,z,2i-1)]}{2}
\]

and

\[
N(x,y,z,1:16) = \frac{[E(x,y,z,2i) - E(x,y,z,2i-1)]}{2},
\]

respectively and where i = 1, ..., 16. and stands for mean and standard deviation.

We thereafter compared the SNR dependence as a function of the spatial position in the magnet by plotting it along a diameter through the phantom volume, as shown in Fig. 3b).

Eddy-current-measurements

In high-angular resolution diffusion experiments high b-values are applied that may introduce residual eddy-current related image distortions that depend on the diffusion weighting of each frame. This could lead to image blurring, erroneous ADC and FA estimates and could affect fiber tract computation in the affected areas. Eddy currents (EC) were measured by acquiring 9 slices in a phantom with 4 b-values (0, 400, 700 and 1000 s/mm²) and 12 directions. An EC constrained registration was performed to obtain the expected shear, scale and translation parameters as a function of the diffusion gradient amplitude (Andersson and Skare, 2002). Thus the shear, scale and translation parameters for a b-value 6400 s/mm⁴ scan could be estimated. The coefficients of the EC constrained registration were analyzed to assess the impact of distortions on the subject study.
In-vivo study

A total of eight healthy volunteers (4 females and 4 males, age = 29 ± 5 years) were included in this study. None of the subjects had a history of neurological or psychiatric disorders. The ethical committee of every hospital involved in this work approved this study and a written informed consent was obtained from each participant.

Data acquisition

All data acquired at the four scanners had a matched protocol as follows. DSI data was acquired using a twice-refocused single-shot echoplanar imaging (EPI) product sequence and the following parameters: TR/TE = 8000/154 ms, voxel size = 2.2 × 2.2 × 3 mm, 44 slices, 256 directions covering half of the q-space with 4 shells and a maximum diffusion weighting with b = 6400 s/mm², and one image acquired without diffusion weighting (b = 0 s/mm²). The twice-refocused diffusion encoding applies a bipolar diffusion encoding scheme to minimize residual eddy-current sensitivity (Roemer et al., 1990). For anatomical reference, 3D T1-weighted MPRAGE volumes (TR/TE = 2300/2.86 ms, FoV = 240 × 256 mm, voxel size = 1 × 1 × 1.2 mm) and 3D T2-weighted SPACE images (TR/TE = 3200/408 ms, FoV = 256 × 256 mm, voxel size = 1 × 1 × 1.2 mm) were acquired for each subject.

The following measurements were performed in order to assess reproducibility:

(i) Intra-scanner reproducibility: all the subjects underwent an MRI scan twice on the same scanner (scanner A), but at different times.

(ii) Reproducibility with twin-scanner settings: all the subjects underwent an additional MRI session on scanner B, which is of the same model as scanner A and is equipped with the same coil hardware and identical reconstruction settings were used as with scanner A.

(iii) Reproducibility with other configurations: five out of the eight subjects were scanned on all four scanners at the three hospitals. Besides the twin scanners A and B, this also included scanners C and D. Scanner C differs from the twin scanners in the employed reconstruction settings as the signals from the different coil elements were combined using the so-called sum-of-square method (Roemer et al., 1990). All other scanners used the conventional adaptive-combine combination of coil signals. Scanner D was equipped with a different RF-coil hardware using a 12-channel head matrix coil (all other scanners were equipped with a 32-channel head coil).

All scanning sessions were carried out within a six month time interval, to minimize variability in the data due to biological changes as sources for obtained reproducibility. To minimize the motion artifacts we fixed the head of the subject with cushions and instructed them to avoid any motion during the 30-minute scan. At scanner B, an additional scanning session was conducted with five subjects using the reconstruction settings of scanner C (sum-of-squares).

Image processing

The intra- and inter-scanner variability of diffusion measurements was assessed using (i) a region-based evaluation of scalar maps and (ii) a tract-based evaluation of scalar maps. These two analysis techniques have also been employed in recent diffusion studies (Bonekamp et al., 2007; Lau and Goodyear, 2007; Vollmar et al., 2010).

The orientation distribution function (ODF) was estimated using the Diffusion Toolkit (www.trackvis.org/dtk). The ODF is evaluated for a set of vectors which are the vertices of a regular polyhedron, the 362 vertex 6-fold geodesated icosahedron, of mean nearest-neighbor separation = 0.16, rad = 9°.

The GFA was directly computed from the DSI raw data. Tuch (2004) defines the GFA as an analog for q-ball imaging of the FA in DTI. The GFA metric is defined as

\[
GFA = \frac{SD(ODF)}{RMS(ODF)}
\]

where SD(ODF) is the standard deviation of the orientation distribution function ODF and RMS(ODF) is its root mean square.

The computation of FA and ADC is based on diffusion weighted data with sufficient SNR. Since the DSI data contains images with weightings of up to b = 6400 s/mm², only a subset of the DSI data is considered for the calculation of the FA and ADC scalar maps. In fact, a DTI dataset was obtained by linear interpolating point on a q-space shell at b = 1000 s/mm² along 64 directions distributed on the unit sphere.

(i) Region-based-evaluation. The obtained FA, ADC, and GFA scalar maps were registered to the MNI space, where the corpus callosum splenium (CCS), cortico-spinal tract (CST) and uncinate fasciculus (UF) were manually delineated by an expert neurologist (S.V., 10 years of expertise) as depicted in Fig. 1. Registration was carried out using the nonlinear registration tool of FSL (www.fmrib.ox.ac.uk/fsl/). The mean value of each scalar map was computed inside every region-of-interest (ROI).

(ii) Tract-based-evaluation. Fiber-tracking was performed using an in-house streamline-based algorithm adapted to work with DSI data. 30 seed points were created inside each voxel in the white matter. Then, from the MPRAGE image of each subject 70 cortical and 8 sub-cortical regions with anatomical landmarks were mapped using FreeSurfer 5.0 (surfer.mrri.ohg.harvard.edu), and subsequently non-linearly registered to diffusion data. Six bundles-of-interest were isolated by selecting pairs of regions of interest known to be connected through these tracts, as shown in Fig. 2. These tracts were identified by an expert neurologist. The mean value of each scalar map was computed along every single fiber of a tract of interest and then averaged.

The variability of the region-based and the tract-based analysis was assessed by means of the coefficient of variation (CV) of the measurements, which is the most commonly used relative measure of reproducibility reported in literature (Bonekamp et al., 2007; Ciccarelli et al., 2003; Heiervang et al., 2006). The CV is defined as the standard deviation of certain measurements divided by the corresponding mean.

\[
CV = \frac{\sigma}{\mu}
\]

We determined the CV for each subject between two scanners and thereafter average the CVs, that is the mean CV that was computed for each subject.

To further investigate any possible correlation between the scanners we also computed the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979). The ICC considers both the within-subject variance arising from measurement error and the variance from the difference between subjects, according to

\[
ICC_{3,1} = \frac{BMS-EMS}{EMS + (K-1)EMS}
\]

where BMS is the between subject mean square (between subject variation), EMS is the error mean square and k is the number of scanners.
Whether we calculate CV or ICC, we always consider only the analysis between a pair of scanners.

**Results**

**Phantom**

As demonstrated in the phantom measurements in Fig. 3a), scanners A, B, and C – all equipped with a 32-channel head array – perform very similarly. Scanner D (equipped with a 12-channel head matrix coil) shows very similar SNR in the center of the coil but lower SNR closer to the proximity of the coil elements.

The eddy-current measurements on the phantom showed that, even though the acquisition protocol was designed to be particularly sensitive to the eddy-currents, eddy currents were negligible across all scanners. As described in Materials and methods we applied an EC model, similar to Andersson and Skare (2002), to estimate the expected translation, scaling and shear for the worst case scenario, b-value = 6400. This showed that errors due to scaling, translation and shear were all equally affected on all sites. The ranges of pixel shift for translation, scale and shear across the sites are: 0.0424–0.42, 0.015–0.3 and 0.048–0.48.

**Scanner A and twin-scanners**

The first column in Fig. 4 reports a summary of the CV of the mean GFA, FA and ADC inside the ROIs and along the considered tracts. Some interesting considerations arise from these comparisons: first, the averaged CV for all scalar maps are in a range of 1%–4.2% for the region-based analysis and below 3% for all computed bundles of the tract-based approach except for the UF. The UF showed the lowest reproducibility, especially for FA and GFA values (CV ~ 3.7–6%). Second, a paired t-test showed no significant differences for both the intra- and inter-scanner comparisons (intra: p>0.17, inter: p>0.06). Third, in general the ADC shows the best reproducibility as compared to FA and GFA, while FA and GFA show very similar reproducibilities.

The second column of Fig. 4 compares the variability of the scalar measures obtained at the two twin-scanners by means of ICC. Values of ICC above 0.7 are considered as measures of high reproducibility (Marenco et al., 2006). From the figure, it appears that both FA and GFA have an ICC above this threshold in the majority of the investigated regions and tracts. No significant difference can be observed among them.

This is however not the case for ADC, where some low ICC values down to 0.3 can be observed for smaller regions, bundles and inter-hemispheric bundles. Finally, the last observation is that intra-
scanner acquisitions showed, on average, slightly better ICC (0.71 (ADC), 0.83 (FA) and 0.86 (GFA)) than the inter-scanner (0.68 (ADC), 0.83 (FA) and 0.85 (GFA)), although this difference was not significant (ADC: p > 0.71, FA: p > 0.95, GFA: p > 0.76).

Different scanners

Fig. 5 summarizes the comparison between scanner A vs itself, A vs C and A vs D concerning the reproducibility of the three considered

![Image](image_url)

**Fig. 3.** a) The log(SNR) of the four scanners along a diameter. b) SNR map of scanner A.

![Image](image_url)

**Fig. 4.** ICC and CV of GFA (1st row), FA (2nd row) and ADC (3rd row) of the inter(green)-and intra(red)-site comparison of scanners A and B, CV is represented as mean ± standard deviation. ROIs: corpus callosum splenium (CCS), cortico-spinal tract (CST) and uncinate fasciculus (UF). Tracts: uncinate fasciculus (UF), corpus callosum body (CCB), corpus callosum splenium (CCS), thalamo-cortical fibers (TCF), fronto-occipital fasciculus (FOF) and arcuate fasciculus (AF).
scalar maps (i.e. ADC, FA and GFA) for the region-based and tract-based analyses. As in the previous section, the CV of the mean GFA, FA and ADC inside the ROIs and along the considered tracts is reported. The worst CV is for the A vs D, 9%–16% for the GFA, 4%–8.9% for the FA and 4.2–10% for the ADC. Here again, the CV for the UF tend to have a consistently higher standard deviation for all three scalar maps, reflecting the relative size of this bundle compared to the rest. A paired t-test showed no significant difference for the A vs D comparison when looking at the GFA and FA maps, but this did not hold for the ADC map where five of the ROIs and tracts showed p<0.02. The CV for A vs C showed more significant different regions and tracts especially in the case of GFA when all ROIs and tracts but the UF were significantly different at p<0.036.

Fig. 6 highlights the relation of the FA vs. the ADC scalar maps for all the four scanners. Mean ADC and FA values are reported in the UF for the region-based analysis and along the CCB bundle for the tract-based analysis. Similar trends (not shown here) hold for all combinations of scalar measure (FA, ADC and GFA) and region/bundle considered. It can be noted that there is a very high similarity of the outcomes from scanners A and B, which are indeed identical imagers. However, FA and ADC measures from scanner D (12-channel coil) are slightly lower and an even more pronounced shift can be observed for scanner C (sum-of-square coil combination). Each color represents a certain scanner and each symbol a specific subject.

As mentioned in the Materials and methods section, five additional subjects were scanned on scanner B with the same reconstruction settings as scanner C (sum-of-square) to confirm the hypothesis that the shift of scanner C in Fig. 6 is indeed due to the different reconstruction method.

Discussion

The intra- and inter-scanner variability of the DSI measurements can be affected by a number of factors, including differences in hardware, protocol settings, the signal-to-noise level, eddy current distortions, field inhomogeneity, subject positioning, head and physiological motions, differences in the reconstruction and post-processing strategies. Some of these contributors can be assessed with phantom scans, i.e. we investigated the SNR and the distortions induced by the eddy current. The observation of the highest SNR with the 32-channel coil in the proximity of the coil elements, which at the same time exceeds the SNR obtained with the 12-channel coil in this area, is expected due to the coil design and is in very good agreement with recent publications (Wiggins et al., 2006). SNR properties are very similar for scanners A, B, and C. However, the 12-channel head array used in scanner D provides a more than two fold reduction of SNR at around 8 cm from the coil center, which at a first glance corresponds to the cortex location in human in-vivo measurements. As SNR differences may introduce variations in the FA, ADC and GFA computation (Burlette et al., 2001; Farrell et al., 2007), i.e. in this investigation that relies on an interpolation of a q-space shell with b=1000 s/mm², it appears advantageous for multi-centric studies to ensure similar SNR across the platforms.

Fig. 6 witnesses that there is a bias shift for the FA and ADC measures obtained for scanners C and D. The 12-channel coil array used in
scanner D provides a lower SNR away from the coil center, which most likely explains the observation of shifted FA and ADC values compared to scanners A and B. The shift in scanner C originates from a different reconstruction setting used for the combination of the signals from different coils. Whereas scanners A, B, and D used an adaptive-combine combination that generally leads to a lower noise contamination (in other words higher SNR), scanner C utilized the so-called sum-of-square coil combination (Roemer et al., 1990) method. This interpretation of the results was confirmed by an additional acquisition on scanner B for a subset of the subjects that used the sum-of-square method for the reconstruction as explained in the Results section.

It has been shown in Walsh et al. (2000) that compared to the sum-of-square technique, the RMS noise level in dark image regions is reduced by as much as $\sqrt{N}$, where N is the number of coils in the array. The sum-of-square method also tends to inadvertently enhance motion and flow artifacts in dark pixels. It is therefore recommended to use adaptive combine when the signal level is comparable to the noise level (choleangiography, DTI/DWI, etc.). In Benner et al. (2006) the tractography yields better results when using adaptive combine compared to the sum-of-square reconstruction.

Summarizing, SNR can be modulated by the field strength, RF-coil hardware, and reconstruction algorithms including coil combination, but also by other parameters such as parallel imaging implementation and others. The results in this study indicate that the computed scalar maps reveal clear sensitivity to the underlying SNR. Thus, it is recommended to closely monitor all hardware and protocol settings of the imaging platforms in a multi-center setting.

Similarly eddy current behavior and field inhomogeneities after shim should be monitored. Our results indicated that at large $b$-values the distortion is expected to be on the order of 0.5 pixels; however, the different eddy current characteristics did not introduce a systematic variability across these scanners. In a multi-centric setting, particular attention of the potential systematic bias degrading the reliability of the diffusion parameters always needs to be considered. If one scanner has significantly different, and large, eddy current characteristics, strategies to correct the distortions need to be implemented. Ideally if time is not a constraint, the residual eddy currents should be measured and corrected through a field map distortion (Truong et al., 2008), alternatively image registration procedures that utilize information from low $b$-value scans to predict the correction of higher $b$-values similar to Andersson and Skare (2002) have shown merit. In addition measuring the field inhomogeneities provides an attractive way to retrospectively correct for susceptibility induced image distortions that may improve subsequent processing steps.

In this study we can notice that GFA and FA carry coherent information regarding the local diffusion anisotropy as their values computed from different region/tracts on different imagers follow comparable trends. This has been confirmed in clinical studies as well (Chiu et al., 2011; Liu et al., 2010; Lo et al., 2011; Tang et al., 2010).

Another factor that has been demonstrated to affect the reproducibility of the DWI results is cardiac pulsation (Pierpaoli et al., 2003), which showed that cardiac gating may provide a better intra- and inter-session reproducibility in some brain regions (Vollmar et al., 2010). Unfortunately, gating has not been widely accepted in clinical studies because of the extended scanning time and occasional instability of gating depending on patient conditions such as arrhythmia. In this study, we decided not to use cardiac gating to represent the most complete information about the underlying diffusion process and the strategy used to combine the signals from each separate coil element. However, if the SNR is matched across the acquisitions, the use of properly designed protocols makes it possible to pool DSI data from different scanners. The results suggest that DSI can be reliably used in clinical research involving different imaging centers, hence increasing the power of the clinical studies and the sensitivity of diffusion MRI to brain pathologies and neurological disorders.

Conclusions

In this study we investigated the intra- and inter-site variability of DSI measurements at four different 3 T scanners. We investigated the diffusion MRI scalar measures (FA, ADC and GFA). We have focused our attention on DSI, as it is a technique capable of providing the most complete information about the underlying diffusion process and allows disentangling more complex fiber architecture such as fiber kissing and crossings. To our knowledge, the reproducibility of DSI has never been studied so far. We have shown that reproducibility at a level as good as the standard clinical DTI acquisitions can be achieved provided that some factors are taken into consideration. The most critical factors in our study affecting the variability of the aforementioned outcomes turned out to be the obtained SNR that is modulated by the number of channels in the receive coil and the strategy used to combine the signals from each separate coil element. However, if the SNR is matched across the acquisitions, the use of properly designed protocols makes it possible to pool DSI data from different scanners. The results suggest that DSI can be reliably used in clinical research involving different imaging centers, hence increasing the power of the clinical studies and the sensitivity of diffusion MRI to brain pathologies and neurological disorders.

Acknowledgments

This research was supported by the Swiss National Science Foundation, SPUM (grant nos 33CM30-124089 and 320030-141165) and by the Centre d’Imagerie BioMédicale (CIBM) of the University of Lausanne (UNIL), the Swiss Federal Institute of Technology Lausanne (EPFL),
References


