NeuroImage 259 (2022) 119439

Contents lists available at ScienceDirect

NeuroImage

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

Cross-site harmonization of multi-shell diffusion MRI measures based on rotational invariant spherical harmonics (RISH)

Alberto De Luca[a,b,*], Suheyla Cetin Karayumak[c], Alexander Leemans[b], Yogesh Rathi[c], Stephan Swinnen[d,e], Jolien Gooijers[d,e], Amanda Clauwaert[d,e], Roald Bahr[f], Stian Bahr Sandmo[f], Nir Sochen[g,h], David Kaufmann[i], Marc Muehlmann[j], Geert-Jan Biessels[a], Inga Koerte[c,k], Ofer Pasternak[c]

[a] Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands
[b] PROVIDI Lab, Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands
[c] Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[d] Movement Control and Neuroplasticity Research Group, KU Leuven, Leuven, Belgium
[e] KU Leuven Brain Institute (LBI), Leuven, Belgium
[f] Oslo Sports Trauma Research Center, Norwegian School of Sport Sciences, Oslo, Norway
[g] Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel
[h] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
[i] Radiology Department, Charite University Hospital, Berlin, Germany
[j] Department of Radiology, Ludwig-Maximilians-Universität, Munich, Germany
[k] cBRAIN, Department of Child and Adolescent Psychiatry, Ludwig-Maximilians-Universität, Munich, Germany

A B S T R A C T

Quantification methods based on the acquisition of diffusion magnetic resonance imaging (dMRI) with multiple diffusion weightings (e.g., multi-shell) are becoming increasingly applied to study the in-vivo brain. Compared to single-shell data for diffusion tensor imaging (DTI), multi-shell data allows to apply more complex models such as diffusion kurtosis imaging (DKI), which attempts to capture both diffusion hindrance and restriction effects, or biophysical models such as NODDI, which attempt to increase specificity by separating biophysical components. Because of the strong dependence of the dMRI signal on the measurement hardware, DKI and NODDI metrics show scanner and site differences, much like other dMRI metrics. These effects limit the implementation of multi-shell approaches in multicenter studies, which are needed to collect large sample sizes for robust analyses. Recently, a post-processing technique based on rotation invariant spherical harmonics (RISH) features was introduced to mitigate cross-scanner differences in DTI metrics. Unlike statistical harmonization methods, which require repeated application to every dMRI metric of choice, RISH harmonization is applied once on the raw data, and can be followed by any analysis. RISH features harmonization has been tested on DTI features but not its generalizability to harmonize multi-shell dMRI. In this work, we investigated whether performing the RISH features harmonization of multi-shell dMRI data removes cross-site differences in DKI and NODDI metrics while retaining longitudinal effects. To this end, 46 subjects underwent a longitudinal (up to 3 time points) two-shell dMRI protocol at 3 imaging sites. DKI and NODDI metrics were derived before and after harmonization and compared both at the whole brain level and at the voxel level. Then, the harmonization effects on cross-sectional and on longitudinal group differences were evaluated. RISH features averaged for each of the 3 sites exhibited prominent between-site differences in the frontal and posterior part of the brain. Statistically significant differences in fractional anisotropy, mean diffusivity and mean kurtosis were observed both at the whole brain and voxel level between all the acquisition sites before harmonization, but not after. The RISH method also proved effective to harmonize NODDI metrics, particularly in white matter. The RISH based harmonization maintained the magnitude and variance of longitudinal changes as compared to the non-harmonized data of all considered metrics. In conclusion, the application of RISH feature based harmonization to multi-shell dMRI data can be used to remove cross-site differences in DKI metrics and NODDI analyses, while retaining inherent relations between longitudinal acquisitions.

Introduction

Diffusion magnetic resonance imaging (dMRI) is an important tool to non-invasively probe the in-vivo organization of the human brain. dMRI metrics such as mean diffusivity (MD) and fractional anisotropy (FA) derived from diffusion tensor imaging (Basser and Pierpaoli, 1996) (DTI) can capture subtle properties of white matter (WM) microstructure, such as changes during neurodevelopment (Girault et al., 2019), aging (Cole, 2020), neurodegeneration (Gyebnár et al., 2018) and in focal insults (Arfanakis et al., 2002).

Most dMRI studies to date are based on a modest sample size, in the order of tens of subjects. However, studies with moderate to low effect sizes (Smith and Nichols, 2018), e.g., imaging-genetics experiments or early pre-symptomatic microstructural changes in traumatic brain injury, require larger sample sizes. The importance of large sample sizes

* Corresponding author.
E-mail address: a.deluca-2@umcutrecht.nl (A. De Luca).

https://doi.org/10.1016/j.neuroimage.2022.119439.
Received 15 February 2022; Received in revised form 23 June 2022; Accepted 30 June 2022
Available online 3 July 2022.

to confirm explorative findings of small sample studies or to study rare conditions is increasingly recognized in neuroimaging (de Luca and Biessels, 2021; Turner et al., 2018). This is also demonstrated by the impact of recent large cohort studies such as the UK Biobank (Allen et al., 2012) or the Human Connectome Project (Van Essen et al., 2013). Achieving large sample sizes in a single acquisition site is, however, challenging. A common alternative is to collect data across multiple sites. Nevertheless, multi-site acquisition has limited usability for dMRI, because of marked inter-site variability. Differences in gradients hardware, receive and transmission coils, magnetic field inhomogeneity and reconstruction software contribute to offsets in the dMRI measurements, introducing complicated and non-linear site differences in dMRI derived metrics, even when identical scanners and acquisition protocols are employed (Mirzaalian et al., 2016).

Recently, a cross-site harmonization method based on rotational invariant spherical harmonics (RISH) (Cetin Karayumak et al., 2019; Mirzaalian et al., 2016) was proposed, among others, to retrospectively harmonize dMRI data collected at multiple sites. In the RISH features harmonization method, data from multiple diffusion directions with the same b-value (single-shell) are deconstructed into a set of spherical harmonics up to a given order. Then, a RISH feature is computed for each even harmonic order as the squared sum of the corresponding harmonic coefficients. The RISH features are calculated for each voxel, or can be averaged across regions of interest. Given dMRI data of matched control groups that have been collected at two different acquisition sites, the method assumes that any group difference in the RISH features is entirely due to scanner-related effects. Accordingly, harmonization is trained by learning the spatially varying mapping of the average RISH features from one site to another, which can be then applied to harmonize data of any given subject. Because the harmonization can be performed at the voxel-level, the RISH method can potentially model bulk inter-scanner differences, but also scanner differences stemming from spatial inhomogeneities that might be due to differences in gradients, coil sensitivity, B0 and B1 fields, among others. Unlike methods such as ComBat (Fortin et al., 2018) that perform harmonization at the statistical inference stage for each derived measure of interest, the RISH-based method harmonizes the dMRI data once as part of the pre-processing. This allows for the subsequent derivation of any harmonized measure of interest and for the application of any analysis method of choice across the harmonized subjects, including whole brain statistics, region of interest based analysis, tract based spatial statistics or voxel-wise analyses.

Thanks to the advent of novel acceleration techniques such as simultaneous multi slice (Setsompop et al., 2018), compressed sensing (Ning et al., 2015), and other recent hardware improvements, the implementation of dMRI protocols beyond the single-shell is becoming increasingly popular even in clinical settings. The most common acquisition is of a set of spherically distributed gradient directions at multiple diffusion weightings (i.e., *multi-shell*). Multi-shell data allows fitting more complex models than DTI, such as higher-order models that explore the non-Gaussian and multi-exponential nature of the dMRI signal, providing complementary and / or additional information to DTI metrics (De Luca, 2021). A popular higher-order model is the diffusion kurtosis imaging (DKI) (Jensen et al., 2005; Jensen and Helpern, 2010) method, which extends the DTI formalism to the second order cumulants. DKI models 3D diffusion processes that deviate from Gaussianity due to, for instance, the presence of membrane restrictions (Fieremans et al., 2011) or intra-voxel incoherent motion effects (De Luca et al., 2017). The non-Gaussian behavior is quantified as excess kurtosis of the diffusion process and has been shown to provide complementary information to DTI. For example, DKI has provided useful information about the aging brain (Falangola et al., 2008), and has also been able to detect pathological changes with higher sensitivity than DTI in patients with Alzheimer's disease (Arab et al., 2018), among others. Another family of methods leveraging multi-shell acquisitions are biophysical models. Such models are often based on assumptions on the tissue microstructure, and aim to quantify metrics that directly relate to specific biophysical processes. A popular example hereof is the "neurite orientation dispersion and density imaging" (NODDI) model, which over time has found widespread application to characterize the healthy and pathologic brain (Calabrese et al., 2017; Colon-Perez et al., 2019; De Luca et al., 2020; Parvathaneni et al., 2017; Wen et al., 2019).

In this work, we investigate whether the RISH-based harmonization method (Cetin Karayumak et al., 2019), which was previously applied to harmonize single-shell dMRI data, could be effectively generalized to multi-shell dMRI. We tested the effect of the multi-site harmonization on metrics such as the mean kurtosis (MK), FA, and MD derived from the DKI fit, and orientation dispersion index (ODI), neurite density index (NDI) and isotropic signal fraction (ISO) maps from NODDI. A recent work (Tax et al., 2019) provided proof-of-concept that this method can be applied to harmonize MK across sites. However, this was demonstrated on a small group of subjects ("traveling heads") scanned on a number of MRI scanners. Here, we study the application of the RISH features harmonization on multi-shell dMRI data from matching groups of healthy controls acquired at 3 different sites at multiple time points (TP). Taking advantage of a longitudinal study design, we aimed to verify the effectiveness of the RISH harmonization procedure at i) removing baseline cross-site differences in DKI and NODDI values while ii) not altering longitudinal changes in metrics over TPs, which are expected to occur in light of our study sample including teenagers between 14 and 16 years, when rapid structural brain changes occur (Das et al., 2017; Falangola et al., 2008). This effect of harmonization on longitudinal changes in DKI and NODDI metrics was compared between the RISH-based and ComBat harmonization approaches.

Methods

MRI acquisitions

The dMRI data used in this study were acquired in three acquisition sites located in Norway (Oslo, S1), Belgium (Leuven, S2) and Germany (Munich, S3). All the sites were equipped with 3T scanners (Philips Healthcare, Best, the Netherlands) equipped with the standard 32 channels head coil. The MRI at S1 was an Ingenia CX equipped with a CDAS spectrometer, and a gradient system with maximum amplitude equal to 40mT/m. The MRI at S2 was an Achieva dStream equipped with DDAS spectrometer, and a gradient system with maximum amplitude equal to 65mT/m. The MRI at S3 was an Ingenia equipped with a CDAS spectrometer, and a gradient system with maximum amplitude equal to 40mT/m. While all scanners were running on software R5.3 at the beginning of the data acquisition process, Site1 and Site2 were respectively updated to R5.4.1 and R5.4, before acquisition was completed.

The dMRI acquisition was performed at a resolution of $2 \times 2 \times 2\text{mm}^3$. In S2 and S3, the sequence was accelerated with multi-band factor 2 and parallel acceleration SENSE 1.5, resulting in echo time 113 ms and repetition time 7.2 s. S1 did not have multi-band functionality, and therefore the repetition time was changed to 12 s and SENSE to 2 to match the resolution of Sites 1 and 3, while achieving the same echo time (113 ms). The dMRI acquisition of all the three sites included the same gradient table which included 20 x $b = 1000\text{s/mm}^2$ and 30 x $b = 2500\text{s/mm}^2$ in addition to 7 non-weighted images. Additionally, 4 non-weighted images with identical imaging parameters but reversed phase encoding (anterior-posterior) were acquired in each site to correct for EPI-related geometrical distortions. Additional shells that included less than 15 gradient directions, required to derive RISH features of order 4 or above, were omitted from this proof-of-concept demonstration.

Study participants

The participants of this study were young male teenagers recruited as part of the longitudinal REPIMPACT study (Koerte et al., 2021). A total of 46 unique subjects were included in this study. Each subject underwent 1 to 3 MRI sessions, for a total of 79 datasets (Table 1).

Table 1
The second column reports the number (#) of subjects acquired at each site and their distribution per number of timepoints (TP) available between brackets (1/2/3). The last 3 columns report the number of datasets available at each site and timepoint, and the corresponding average age ± standard deviation (std) in years between brackets.

Site	Subjects # (1/2/3 TP)	TP1 # (age ± std)	TP2 # (age ± std)	TP3 # (age ± std)
S1	28 (5/9/14)	22 (14.7 ± 0.6)	22 (15.5 ± 0.6)	21 (15.8 ± 0.6)
S2	17 (4/4/9)	15 (14.5 ± 0.5)	11 (15.2 ± 0.6)	13 (15.5 ± 0.6)
S3	7 (1/1/5)	5 (14.8 ± 0.7)	6 (15.3 ± 0.7)	7 (15.4 ± 0.7)

Ethical approvals were obtained from the ethical review boards of the South-Eastern Norwegian Regional Committee for Research Ethics (S1), KU Leuven (S2) and LMU Munich (S3), in addition to written consent from the legal representatives of the participants.

Data processing and harmonization

The dMRI data were corrected for signal drift (Vos et al., 2016) using ExploreDTI (Leemans et al., 2009) v4.8.6. A brain mask was derived for each subject using FSL BET (Smith, 2002). To remove EPI distortions, FSL TOPUP (Andersson et al., 2003) was applied on the mask, together with the non-weighted images of both phase encodings. Subsequently, FSL EDDY (Andersson and Sotiropoulos, 2016) 5.11 was used to correct for subject motion, eddy currents and EPI distortions in a single step.

The data corresponding to the $b = 1000$s/mm^2 shell (20 gradient directions) and to the $b = 2500$s/mm^2 (30 gradient directions) were fitted with spherical harmonics of order 4 (L4) and 6 (L6), respectively. To perform the harmonization, S1 was chosen as a reference site, since it included the largest number of subjects. For each site to be harmonized, a study-specific template was generated with ANTS (Avants et al., 2014) using selected training subjects of the reference and other site. Specifically, the study template was created following the script "antsMultivariateTemplateConstruction2.sh" of ANTS (https://github.com/ANTsX/ANTs), running an iterative registration procedure based on the FA and L0 images, using the b-spline SyN transformation, and cross correlation as a cost function. The data of the other sites were transformed to RISH features and transformed to the common space to derive the average scale maps needed to remove site effects. The scaling maps were back-projected to the individual space of each subject and used to harmonize its dMRI data. The dMRI data of S1 were also transformed to RISH features but without any scaling of the RISH features. This step was performed to ensure consistency between harmonized and not harmonized data, given that the spherical decomposition might remove signal components in higher order than L4 or L6.

The training of the harmonization between S1 and S2 was performed using TP1 and TP2 of a subgroup of age-matched subjects for both sites (14 unique subjects at TP1 + 1 repeated scan at TP2 for a total of 15 datasets for both sites). For the harmonization of S3, the same subjects from S1 were used as reference. Due to the lower number of participants available at this site, scans obtained at multiple TPs at S3 were used for training harmonization. The training set of S3 consisted of 11 datasets from 7 unique subjects. Subsequently, the DKI model was fit to the data before and after harmonization using FSL DTIFIT and the ordinary least squares approach. This resulted in FA and MD maps, corrected for Kurtosis effects, and MK maps. Physiologically implausible MK values, i.e. MK < 0 or MK < 3 were set to 0. To derive NODDI metrics, we employed the "microstructure diffusion toolbox" (Harms, 2017) (https://github.com/robbert-harms/MDT) running on Python 3.8, and CUDA 10. The fit procedure was accelerated to less than 1 min per dataset by employing a Titan XP GPU kindly donated to ADL from NVIDIA Corporation. From the NODDI fit, we considered the ODI, NDI and ISO maps.

Statistical analyses

The harmonization performance in S2 and S3 were analyzed by creating a study specific template with the FSL TBSS (Smith et al., 2006) registration pipeline. Briefly, the FA maps of each subject were registered to a common space using a combination of FSL "FLIRT" and "FNIRT", then MD, MK, ODI, NDI and ISO were co-registered accordingly. Subsequent analyses included data of subjects not involved in the harmonization step, allowing to evaluate the generalizability of the RISH method beyond the unseen data. To qualitatively verify the effectiveness of the RISH harmonization procedure on the complete dataset, the average RISH features of S1, S2 and S3 were derived for the two shells individually. Next, the scaling between the RISH features of the harmonized and reference sites were evaluated before and after harmonization, and their average value evaluated in a white matter mask (defined as FA ≥ 0.2). To evaluate the similarity between the RISH scaling determined for the two data shells ($b = 1000$, 2500s/mm^2), the Pearson's correlation between the corresponding scaling was determined.

The effectiveness of the harmonization procedure at removing inter-site biases in FA, MD,MK, ODI and NDI was visualized with boxplots of their average value in the whole brain mask. Two-sided t-tests were used to assess the significance of differences in DKI and NODDI metrics between subjects acquired at corresponding TPs at S1, S2 and S3 before and after harmonization. The t-tests were considered significant when p-values < 0.05 were observed (uncorrected). This analysis was also repeated by considering the white and gray matter separately, and can be found in the Supporting Information.

Subsequently, voxel-level differences in FA, MD, MK, ODI and NDI between corresponding TPs of S1 and S2/S3 were assessed by means of two-sided t-tests with FSL randomize using 5000 permutations and the threshold-free cluster enhancement correction for multiple comparisons. This analysis was not performed for the ISO map because it assumed very low values in the brain parenchyma.

To investigate whether dMRI harmonization affects the magnitude of longitudinal changes in DKI metrics, data of S2 was registered to MNI using the same transformations derived to perform voxel-wise statistics. Voxel-wise differences between corresponding scans at TP2 and TP1 were computed, averaged across all subjects of S2 and smoothed with a 5-voxels moving average to mitigate residual mis-registrations. Subsequently, the longitudinal changes of FA, MD, MK, ODI and NDI before and after harmonization were compared by means of the Cohen's d. In addition to the RISH features harmonization, in this step we also considered an additional harmonization method based on post-hoc statistics (Combat (Fortin et al., 2018)).

Results

The RISH features were computed independently for the data at $b = 1000$s/mm^2 and $b = 2500$s/mm^2. Fig. 1 shows the average RISH features of the data acquired in each site with $b = 1000$s/mm^2 before and after harmonization and their relative scaling against the reference site, S1. Before harmonization, the average cross-site scaling in white matter is in the interval 0.99 – 1.02 for S2, and 0.82 – 0.93 for S3. Differences up to ±70% can be observed between S2 and S1 in L0 in the ventricles and the medial temporal area, in L2 at the interface between gray matter

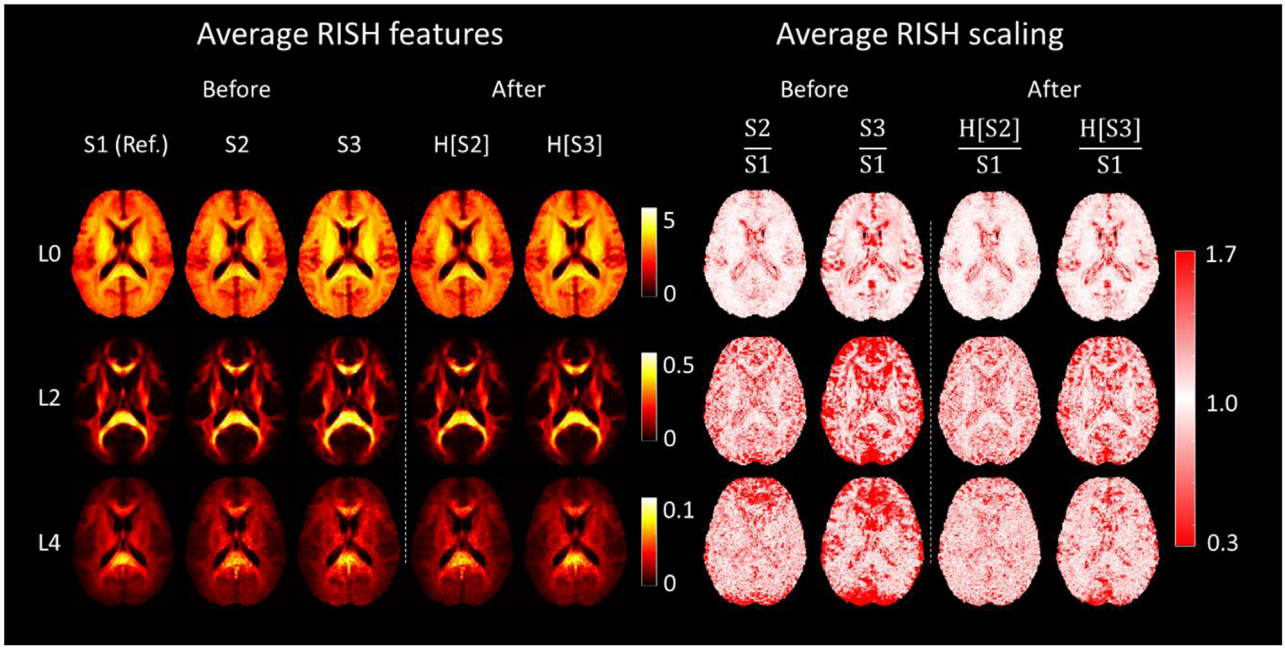

Fig. 1. Harmonization of data at $b = 1000s/mm^2$. Images on the left show the RISH features of all three sites before and after harmonization (H[]). Images on the right side show the voxel-wise scaling between the RISH features of the S2 or S3 site and the reference site, S1. A value of 1 (white color) indicates being identical to the reference site. Before harmonization, large scaling can be observed for L0 in periventricular regions, and in L2 and L4 in frontal and occipital regions.

(GM) and cerebrospinal fluid (CSF), and in L4 in the inferior frontal lobe and occipital lobe. Similar patterns were observed also between S3 and S1 but with larger magnitude. After harmonization, the average cross-site scaling in white matter remains close to 1 for S2 (1.01 – 1.04), and is consistently closer to 1 for S3 (0.94 – 1.00), and most scaling differences in RISH features were removed, especially for L0 and L4. Residual large differences can be spotted in the average L2 ratio maps, but mostly in GM and CSF areas. When looking at whole brain histograms of the RISH features corresponding to $b = 1000s/mm^2$, which are reported in Supporting Information Figure S1, it can be clearly observed that the distributions of RISH features across-sites are more similar after harmonization, albeit some residual differences can be observed in L0 and L4 of S3.

The results of the same comparison for the shell at $b = 2500s/mm^2$ are visualized in Fig. 2. Before harmonization, inter-site scaling patterns that are similar to those in Fig. 1 can be observed, with average values in WM in the interval 1.00 – 1.05 for S2, and 0.89 – 0.97 for S3. In comparison to the results shown in Fig. 1, the harmonization procedure effectively removed most of the inter-site differences in RISH features including L6, despite suboptimal results in L2 for S3 in CSF and GM regions. After harmonization, the average cross-site scaling in white matter at TP1 is close to 1 (1.03–1.04 for S2, 0.98–1.07 for S3). Average histograms of RISH features at $b = 2500s/mm2$, shown in Supporting Information Figure S2, confirm that the distribution shape and average values of RISH features across-sites are closer after harmonization.

Overall, very similar patterns were observed when comparing the scaling of the RISH features of the $b = 1000s/mm^2$ shell to those of the data shell at $b = 2500s/mm^2$. The Pearson correlation coefficients between the RISH features scaling of the $b = 2500s/mm^2$ shell and of the $b = 1000s/mm^2$ shell of S2 were 0.71 for L0, 0.71 for L2 and 0.74 for L4, respectively. For S3, Pearson correlation coefficients were 0.72 for L0, 0.85 for L2 and 0.86 for L4, respectively.

Having established that the individual shells are effectively harmonized using the RISH method, we investigated how harmonization affects the comparability of DKI and NODDI metrics across sites. Fig. 3 shows the FA, MD, MK, ODI and NDI values averaged across a whole brain mask before and after harmonization at each of the TPs. Significant differences were observed before harmonization for all TPs in FA, MD,MK and ODI between S1 and S2 ($p \leq 0.005$) and between S1 and S3 ($p \leq 0.005$), and NDI at TP2 ($p \leq 0.05$). After harmonization, most statistical differences were no longer significant ($p \geq 0.05$) except for residual differences at TP1 between S1 and S3 for FA and MD, although, as the figures show, the harmonized values were closer to the values of S1 than the non-harmonized values. On the contrary, harmonization seemed less effective on NDI at TP3, for which differences between S1 and S2/S3 were only observed after harmonization ($p \leq 0.05$). When looking at WM and GM separately (Supporting Information Fig. S3 and S4), significant differences in both DKI and NODDI metrics were successfully removed in WM, whereas residual differences can be observed in GM.

The results of the voxel-wise comparison of FA, MD, MK, ODI and NDI between S1 and S2 are shown in Fig. 4. Before harmonization, FA values of S1 were statistically lower than that of S2 in several clusters throughout the brain at all TPs, both in GM and in WM regions such as the corpus callosum and the internal capsule. A cluster with lower MD values of S1 was observed in the left internal capsule at TP1 only. Regarding MK, larger values of S1 were observed in the left corticospinal tract and temporal WM. Differences were also observed in NODDI metrics, particularly lower values of ODI and NDI for S1 as compared to S2 at TP1, especially in the right corticospinal tract and temporal WM. After harmonization, differences in FA and MD were completely removed for TP1 and TP2, and reduced by 98.7% and 97.5%, respectively, at TP3. The number of voxels exhibiting statistically different MK values was reduced after harmonization by 100% at TP1, 79.9% at TP2, and 91.9% at TP3. Regarding NODDI metrics, the best results were observed for NDI, for which no significant voxels were observed after harmonization. When looking at ODI, originally observed significant results (S2>S1) were completely removed from harmonization at all TPs. However, the opposite effect (S1>S2) could be observed at TP2 and TP3, although mostly in the corpus callosum next to the ventricles, or in areas affected by susceptibility artefacts.

Fig. 5 reports the results of the same voxel-wise analysis, only now comparing between S1 and S3. Before harmonization, statistically lower FA values, higher MD values, higher MK values, higher ODI values and lower NDI values of S1 were observed in 27.2%, 17.3%, 6.3%, 22.7%

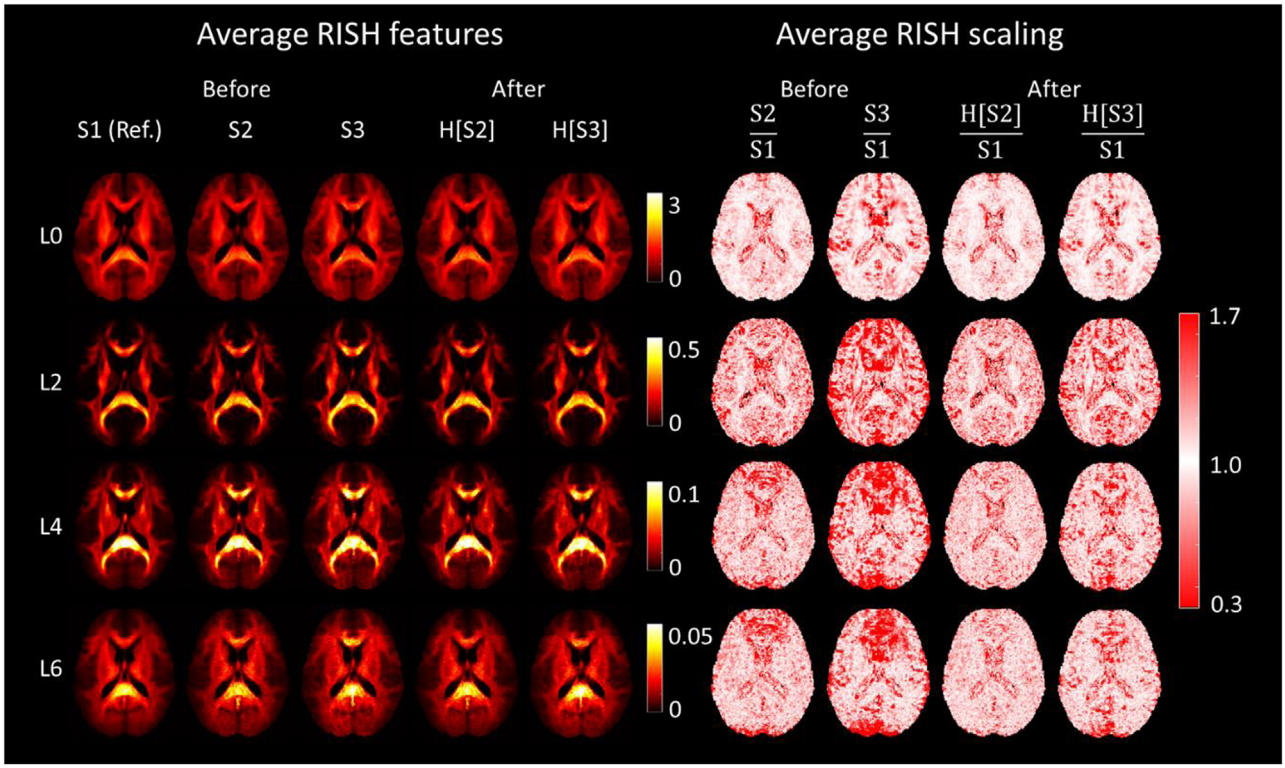

Fig. 2. Harmonization of data at $b = 2500s/mm^2$. Images on the left show the RISH features of all three sites before and after harmonization (H[]). Images on the right side show the voxel-wise scaling between the RISH features of the S2 or S3 site and the reference site. In analogy to Fig. 1, harmonization effectively removes the large scaling observed in frontal and occipital regions of both S2 and S3.

and 2.2% of the brain voxels at TP1, 25.2%, 17.9%, 10.3%, 16.9% and 4.9% at TP2, and 27.4%, 18.0%, 15.0%, 24.7% and 3.9% at TP3, respectively. Most differences in FA were located in WM in the superior parietal lobe, corticospinal tract, frontal and occipital WM, as well as at the WM/GM interface. Regarding MD, most differences were observed in the corticospinal tract and in both frontal and occipital WM. After harmonization, most of the statistically significant differences in FA, MD, MK, ODI and NDI disappeared, reducing the percentage of significantly different brain voxels at TP1/TP2/TP3 to 0.1%/0.1%/0.1% for FA, 0.2%/0.2%/0.2% for MD, 0.1%/0.4%/0.7% for MK, 0.9%/0.2%/0.4% for ODI, and 0.1%/0.2%/0.1% respectively. On the other hand, harmonization seemed to introduce some differences in opposite direction in ODI at TP2 (0.5% of brain voxels) and TP3 (0.1% of brain voxels), and NDI at TP3 (2.7% of brain voxels).

Taking advantage of the presence of multiple TPs in the acquired dataset, we investigated the impact of the RISH based, and ComBat harmonization approaches on longitudinal changes of FA, MD, MK, ODI and NDI of S2 at the voxel level (Fig. 6). Spatially varying patterns of both increase and decrease in all considered metrics were observed before harmonization, and remained mostly unchanged after harmonization with both the RISH and ComBat methods as shown in Table 2.

Table 2
The effect of RISH and ComBat harmonization on the average FA, MD,MK, ODI and NDI values of S2 ± standard deviation at TP1 and TP2 determined in a WM mask (FA≥0.2). The third and fourth columns show the relative percentage change between TP2 and TP1 (ΔTP) and the corresponding effect size (Cohen's d).

	TP1	TP2	ΔTP [%]	Cohen's d
FA				
Original	0.298 ± 0.023	0.290 ± 0.021	−2.7	−0.084 ± 0.333
RISH	0.280 ± 0.012	0.273 ± 0.013	−2.5	−0.085 ± 0.341
ComBat	0.287 ± 0.012	0.279 ± 0.012	−2.8	−0.088 ± 0.346
MD				
Original	8.87 ± 0.04	8.97 ± 0.04	1.1	0.042 ± 0.348
RISH	9.00 ± 0.02	9.12 ± 0.03	1.3	0.052 ± 0.366
ComBat	8.99 ± 0.02	9.10 ± 0.03	1.2	0.044 ± 0.367
MK				
Original	0.82 ± 0.03	0.83 ± 0.02	1.2	0.064 ± 0.359
RISH	0.83 ± 0.02	0.84 ± 0.02	1.2	0.058 ± 0.363
ComBat	0.83 ± 0.02	0.84 ± 0.02	1.2	0.067 ± 0.368
ODI				
Original	0.26 ± 0.01	0.27 ± 0.01	3.3	0.102 ± 0.326
RISH	0.27 ± 0.01	0.28 ± 0.01	3.5	0.108 ± 0.329
ComBat	0.27 ± 0.01	0.28 ± 0.01	3.3	0.106 ± 0.335
NDI				
Original	0.58 ± 0.02	0.59 ± 0.02	1.8	0.098 ± 0.345
RISH	0.58 ± 0.02	0.59 ± 0.02	2.1	0.108 ± 0.356
ComBat	0.58 ± 0.02	0.59 ± 0.02	1.7	0.101 ± 0.357

Discussion

In this manuscript, we investigated the effectiveness of the recently proposed RISH features based harmonization method (Cetin Karayumak et al., 2019) when applied to a multi-shell dMRI acquisition scheme. By evaluating DKI metrics before and after harmonization, we show that the harmonization of individual shells based on the RISH method effectively harmonizes MD, FA and MK across datasets acquired in 3 different sites. Furthermore, the method seems also promising to harmonize NODDI metrics across sites, although less effectively than DKI metrics. We also demonstrate that harmonization preserves the magnitude and variance of changes in DKI and NODDI metrics between timepoints.

The results of this manuscript have several implications for the design and analysis of studies based on multi-site dMRI acquisitions, which are becoming increasingly important to perform robust inference or test biological hypotheses with subtle effects on the diffusion metrics. Our findings support previous studies suggesting that both DKI and DTI met-

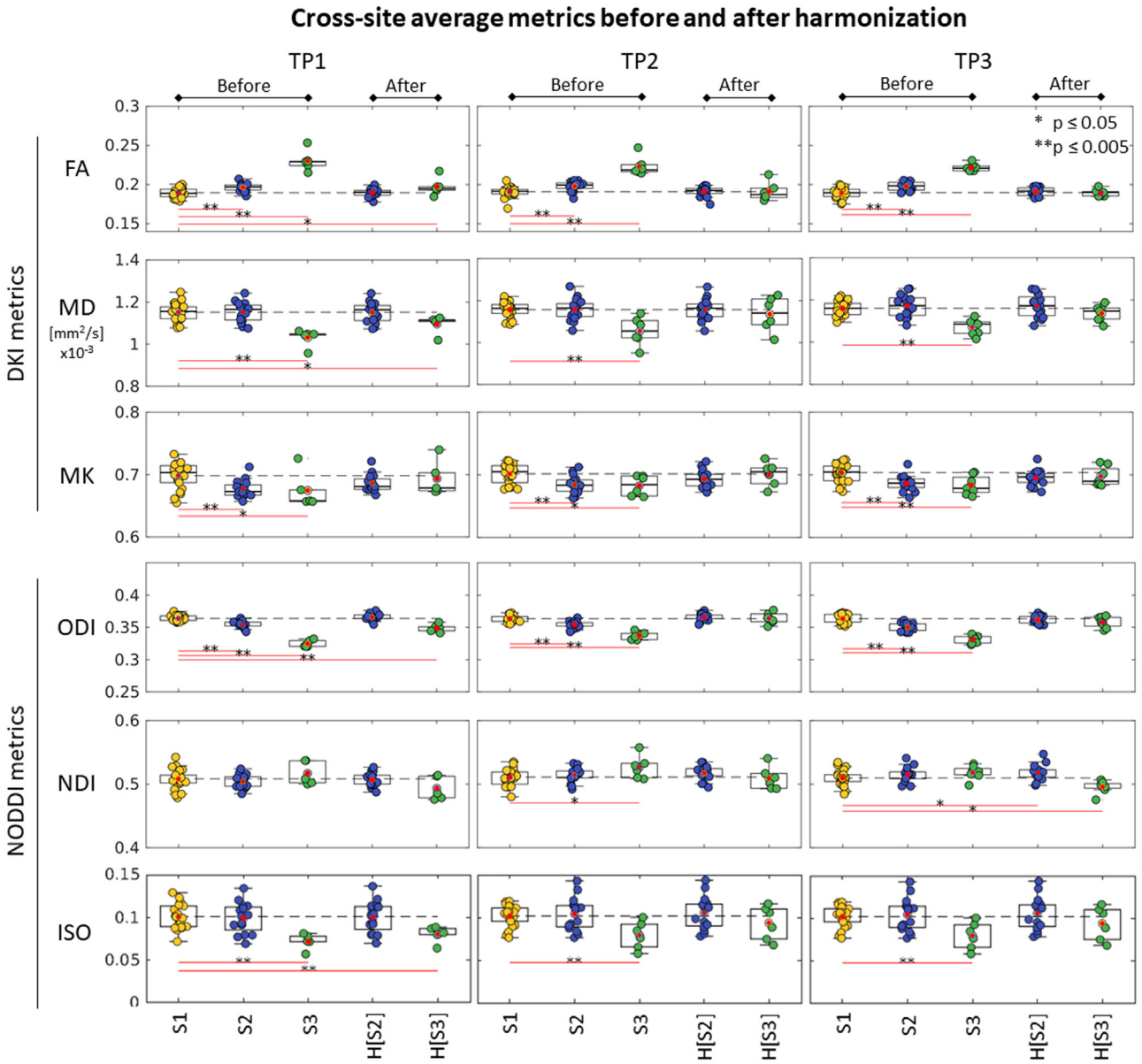

Fig. 3. The boxplots of fractional anisotropy (FA), mean diffusivity (MD), mean kurtosis (MK), orientation dispersion index (ODI) neurite density index (NDI) and isotopic volume fraction (ISO) for S1, S2 and S3 at TP1, TP2 and TP3 before and after harmonization (H[S2] and H[S3]). The black dashed line indicates the average value of S1 for reference. Yellow, blue and green dots represent the average value of the considered diffusion metrics for S1, S2 and S3, respectively. The red dots indicate the average value of each site. The asterisks indicate significant differences between sites: $^{*} = p \leq 0.05$; $^{**} = p \leq 0.005$. After harmonization, most significant cross-site differences are effectively removed, except for small residual differences for FA and MD of S3 at TP1, ODI at TP1 and NDI at TP3.

rics exhibit moderate to large inter-site differences. Here, we add that site differences in metrics are significant even when using MRI scanners from the same vendor and with an almost identical acquisition protocol, and that cross-site differences also affect NODDI quantifications. We found differences in scaling up to more than 70% between the RISH features of S2/S3 and S1, especially in the frontal and posterior region of the brain, and at tissue interfaces (Fig. 1 and Fig. 2). This pattern of differences, which likely originates from a combination of site-dependent coil sensitivity properties and inhomogeneities in the receive and transmit field, hardware differences in gradient systems and / or spectrometers, causes regional differences in DKI metrics across sites (Fig. 4, Fig. 5). Such differences are likely to confound whole brain analyses if not accounted for with a voxel-based harmonization method. In a voxel-wise analysis, inter-site differences might be taken into account as an additional explicit regressor, but will inevitably affect statistical power. Interestingly, the RISH scaling factors determined for $b = 1000\text{s/mm}^2$ (Fig. 1) and $b = 2500\text{s/mm}^2$ (Fig. 2) exhibited remarkable similarity with Pearson's correlation values up to 0.84, supporting the idea that the RISH-based method effectively captures acquisition-hardware related differences. In practice, our result suggests that the RISH harmonization method might be beneficial not only for retrospective studies, but also for prospective studies where residual inter-scanner differences might manifest despite harmonized acquisition protocols. For this reason, we believe this method is relevant to the analysis of data acquired in large multi-site cohorts as the UK BioBank or ADNI, among others.

The RISH-based harmonization method effectively addresses the site-dependence of multi-shell dMRI, reducing the inter-site differences in RISH features between 3 sites (Fig. 1, Fig. 2). Before harmonization, cross-site differences in RISH features increased with the considered RISH order, being minimal with L0 and maximal with L6. The higher order RISH features capture high frequency fluctuations of the signal, such as diffusion anisotropy and noise. The larger site differences in L6 might

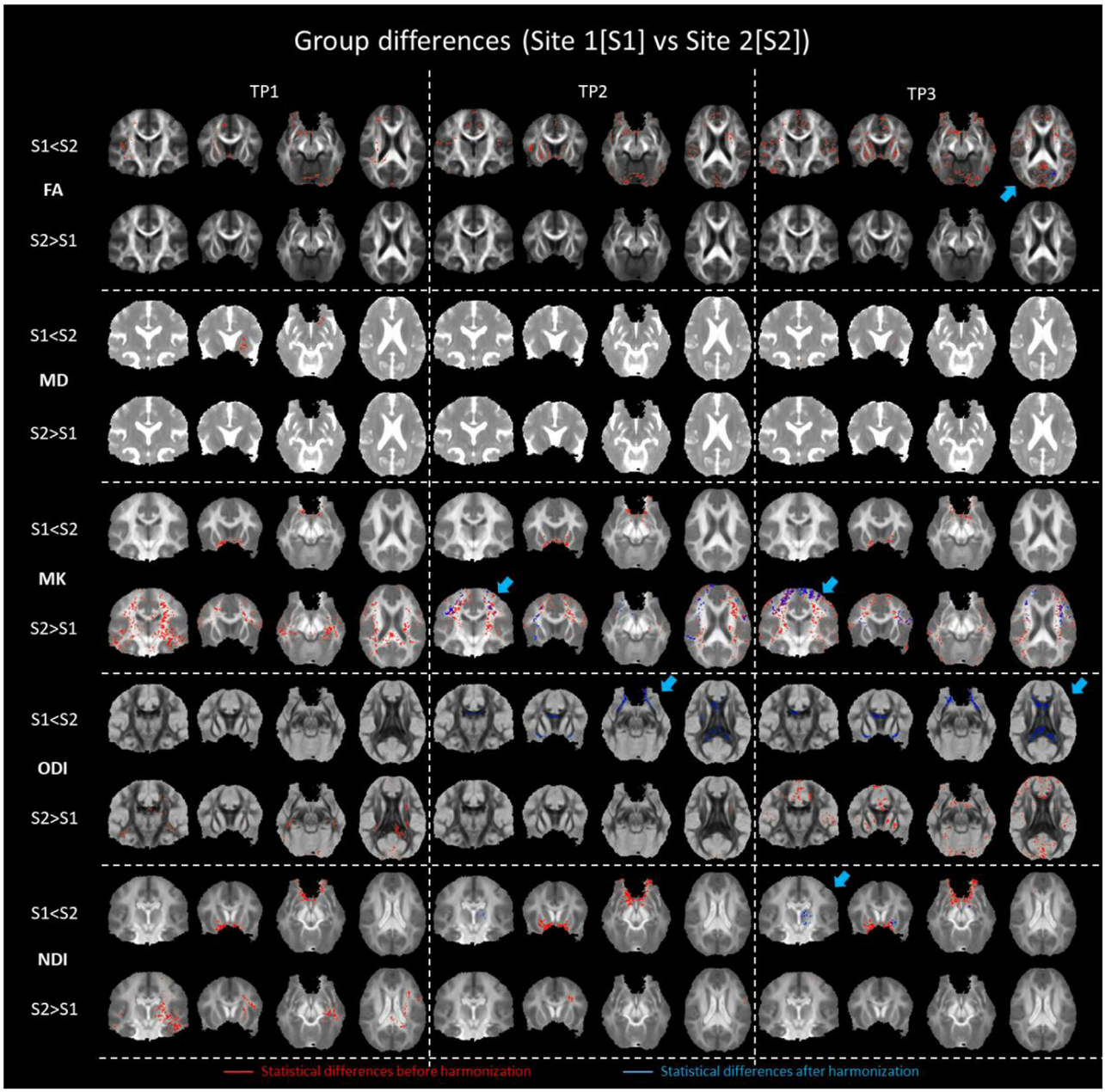

Fig. 4. Voxel-wise analysis of group differences between S1 and S2, before and after harmonization. The different rows present voxels with significant site-difference ($p<0.05$ in a voxel-wise t-test with threshold free cluster correction) when comparing FA, MD, MK, ODI and NDI of S1 to those of S2 before harmonization (in red), and after harmonization (in blue). After harmonization, most significant cross-site differences are removed. Clusters of residual significant voxels (highlighted with blue arrows) can be observed for FA at TP3 in the occipital gray matter, for MK at TP2 and TP3 mostly in the centrum semi-ovale and temporal WM, and for ODI mostly in the ventricles and in temporal WM.

therefore suggest cross-site differences to be driven by differences in diffusion sensitization (e.g., diffusion time) due to, for example, different gradient hardware. This may lead to different sensitivity to diffusion anisotropy and restriction effects, in addition to cross-site differences in effective signal to noise ratio. Integrating denoising strategies in the data pre-processing may provide further insights into the relative contributions of noise to cross-site differences, and deserves further attention in future works as a potential means to remove some of the site differences prior to harmonization.

The RISH method can be considered as part of the dMRI pre-processing steps and only needs to be applied once for each diffusion weighted shell. The RISH method learns cross-site differences only using data of matched controls, and can then be used to harmonize data of target groups (e.g., patients) without any further requirements on the size, demographics or clinical characterization of the target group. This is an important advantage compared to other multi-site harmonization approaches such as meta-analysis or ComBat, which must be applied separately for every statistical inference and for every dMRI measure. Moreover, these methods are sensitive to differences (e.g., demographic) between groups of interest (de Brito Robalo et al., 2021) (e.g., patients vs controls). Unlike the RISH based technique, such methods harmonize each dMRI metric independently not considering the intrinsic mathematical relation between different dMRI metrics. These methods also do not account for spatially varying inter-site differences but rather for bulk effects only, and can be technically challenging to apply at the voxel level in large datasets due to large memory requirements. Another point

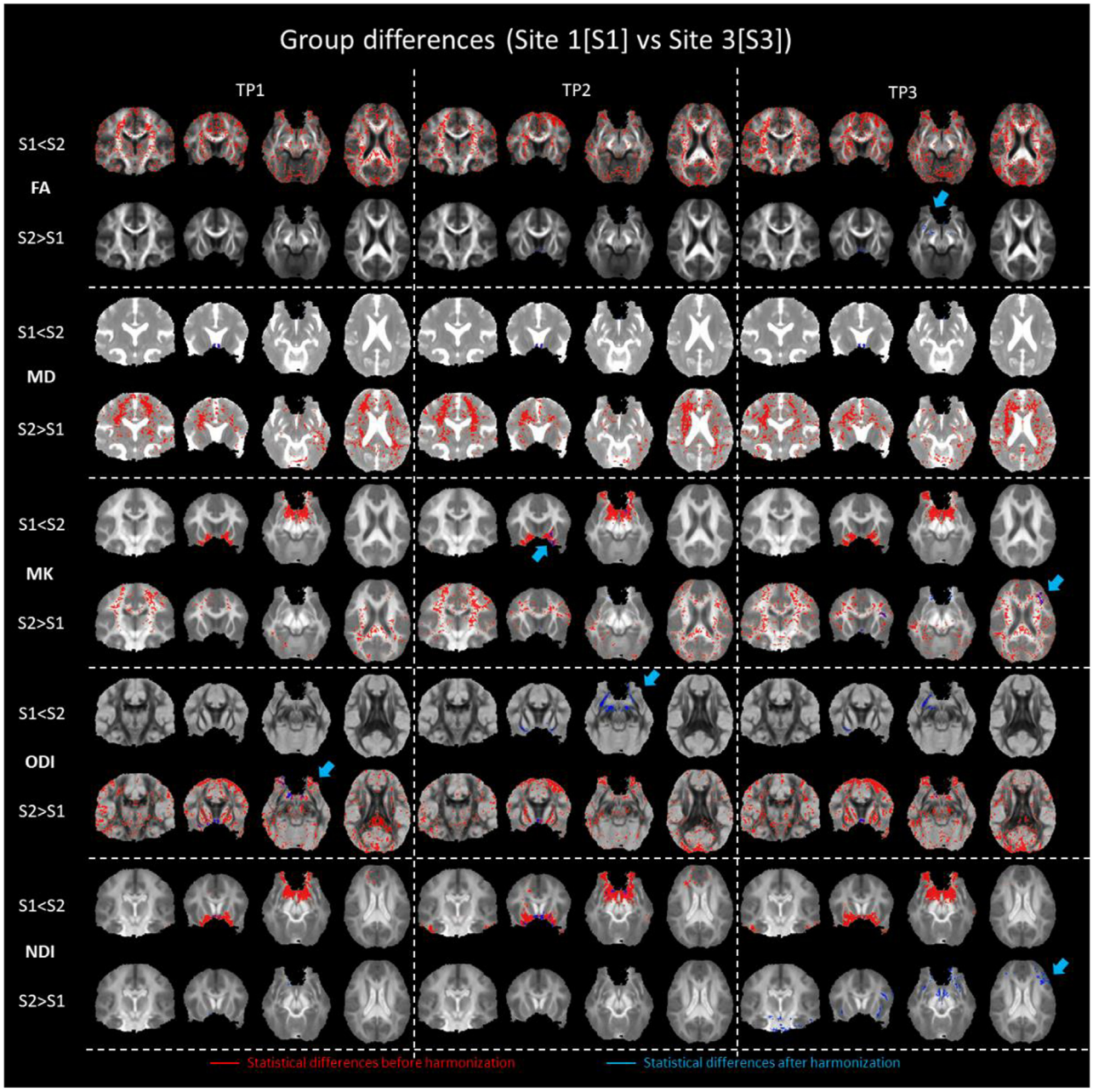

Fig. 5. Voxel-wise analysis comparing S1 and S3 before and after harmonization. The different rows present voxels with significant site-difference ($p < 0.05$ in a voxel-wise *t*-test with threshold free cluster correction) when comparing the FA, MD, MK, ODI and NDI of S1 to those of S3 before harmonization in red, and after harmonization in blue. After harmonization, most significant cross-site differences are removed. Clusters of significant voxels can be observed for FA at TP3 in the bilateral hippocampi and temporal GM, for MK at TP3 in the left frontal WM, for ODI in temporal WM, and for NDI in frontal and parietal WM. Residual differences are highlighted with blue arrows.

to be considered when comparing harmonization methods, is that statistical harmonization methods need to be reapplied every time a new datapoint is added or removed from a study. Should a new subject be added to (or discarded from) the study, the DKI metrics of the remaining subjects would not change with the RISH-based harmonization, whereas they would require recalculation and would change with ComBat to calculate the new site average, which is then matched across sites.

An important desired feature of any harmonization method is that it should not alter the relation between the collected dMRI data and potential effects of interest. To date, most studies on dMRI harmonization focused on showing the removal of inter-site biases in DTI metrics (Cetin Karayumak et al., 2019; Dewey et al., 2019; Mirzaalian et al., 2016; Ning et al., 2020; Pomponio et al., 2020; St-Jean et al., 2020), but this does not guarantee that the abovementioned condition is fulfilled. Recently, we have shown that the RISH method can effectively harmonize data of patients with cerebral small vessel disease while not altering intra-site differences between patients and matched controls (de Brito Robalo et al., 2021). The design of this study offers an opportunity to investigate this aspect in the context of multi-shell harmonization, because it involves participants acquired longitudinally in a lifespan characterized by rapid changes in DKI metrics. In Fig. 6 and Table 2, we reported changes in DKI metrics between consecutive timepoints of the data acquired in S2. With the non-harmonized data, we observe that TP2 is characterized by higher MK values than TP1 in the

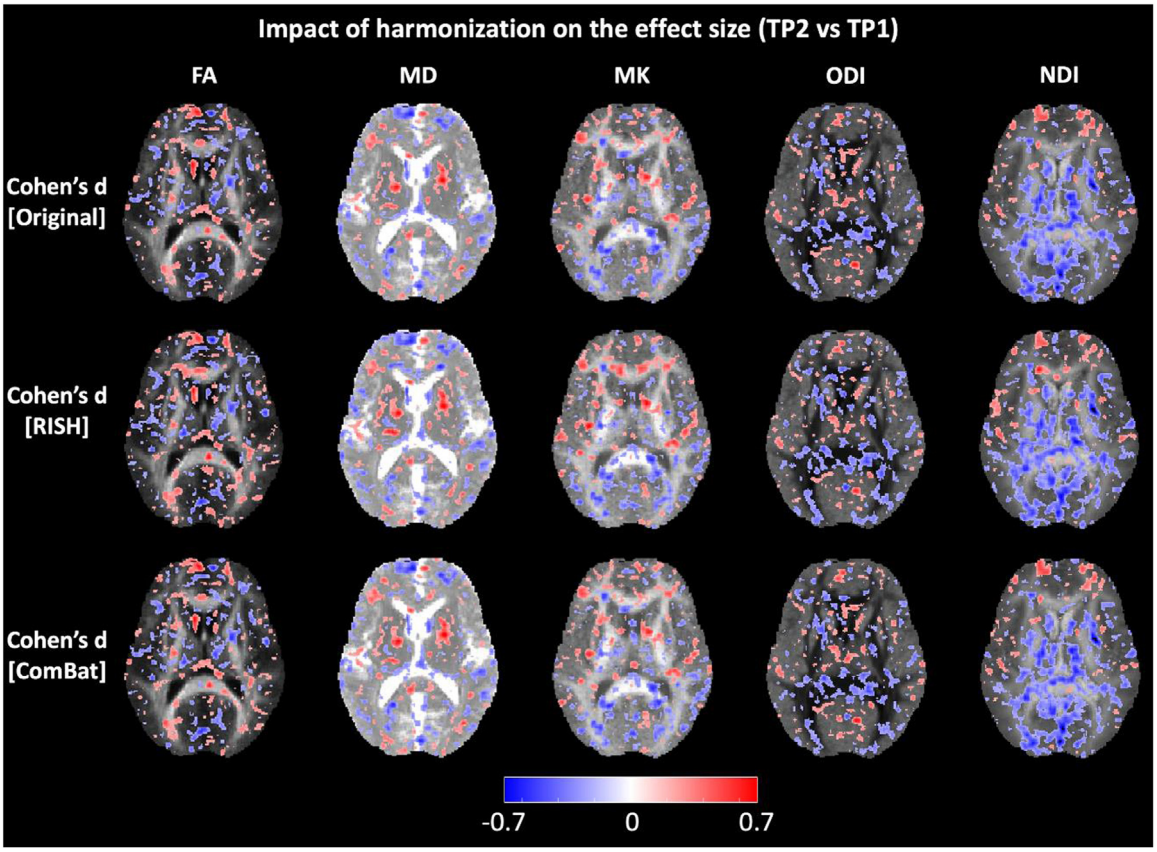

Fig. 6. The average FA, MD, MK, ODI and NDI maps for S2 at TP1 and the overlay of their effect sizes when comparing TP2 to TP1 before harmonization (Original, first row), after RISH harmonization (second row) and after ComBat harmonization (third row). The effect size was determined voxel-wise, smoothed and only effects with absolute value above 0.2 were shown. Both harmonization methods performed similarly and did not generally influence the patterns of longitudinal changes in diffusion metrics. Some differences between the RISH method and data before harmonization can be observed for MK in the frontal part of the corpus callosum.

corpus callosum and thalamus, and lower MK values in the internal capsule and frontal WM. Interestingly, we observed opposite changes in FA (regardless of harmonization or site effects), which is different from previous literature (Lebel et al., 2008; Tamnes et al., 2018). This discrepancy might be explained by the fact that FA derived with the DKI fit is sensitive to slightly different physiological mechanisms than the FA derived with the DTI fit[36]. The harmonization with the tested RISH-based method did not alter the sign and patterns of longitudinal intra-subject changes, which demonstrates preservation of the longitudinal relations between DKI metrics and biological effects. In this analysis, the RISH-based method performed remarkably similar to ComBat. Of note, however, the first was trained on a small subset of subjects mostly from TP1, whereas the latter used all available data from all TPs for harmonization. Some changes of longitudinal effects can be observed in the anterior part of the corpus callosum for both FA and MK, which suggests a potential effect of susceptibility artefacts on the performance of the RISH-based method, although on average this effect can be deemed minor (Table 2).

The harmonization of individual dMRI shells with the RISH method has been previously proven to harmonize DTI metrics when fitting each shell separately[15]. The method has also been shown promising to harmonize DKI metrics derived from a multi-shell fit[13], but only in a group of subjects repeatedly acquired with different MRI scanners. Conversely, in this work we investigated this concept on a broader dataset including different subjects acquired in 3 sites with very similar dMRI protocols (except for the multi-band and SENSE acceleration), which is arguably a more challenging task and closer to expected application. Furthermore, we evaluated the effectiveness of the RISH method to also harmonize metrics derived with the popular NODDI model. On average, cross-site differences are effectively removed by the RISH harmonization method, as shown in Fig. 3. This further generalizes at the voxel level, as shown in Fig. 4 and Fig. 5, although some clusters of residual statistical different voxels can be observed after harmonization for the MK of S2 at TP2 and TP3, and for the MK of S3 at TP3, as well as for ODI of S2 at TP2 and TP3, and NDI of S3 at TP3. Given that TP2 and TP3 were not included in the training of the harmonization of S2, and taking into account the relatively long time between TP1 and TP3 (approximately one year), this result supports that the RISH training can be performed only once when the aim is to derive DKI metrics, provided no major hardware changes occurred, posing the basis for large multi-site data pooling to perform robust statistics of DKI metrics. At the same time, the performance of the method seems to degrade when comparing results at TP1 with TP3, which might be due to subtle changes in the acquisition hardware in the one year between the scanning sessions (e.g., software updates, recalibrations). This seems particularly the case for the NODDI model, and might be due to increased sensitivity to such hardware changes, or to higher scan-rescan variability of NODDI metrics as compared to DKI (Chung et al., 2016; Konieczny et al., 2021; Lehmann et al., 2021), perhaps due to its reliance on a relatively complex fit procedure and on fit constraints that might not be appropriate for all brain tissues (Lampinen et al., 2019). Notably, residual differences in NODDI metrics post-harmonization were mostly observed for the site with the least training subjects and the lowest data quality (S3), and might thus be more the result of limitations in study design than of the RISH method itself. Accordingly, future work should further investigate whether harmonization needs to be trained only once for each site, or whether the training should be periodically repeated to account for possible changes of data quality over time. An alternative explanation for the observed residual differences would be that subjects from differ-

ent sites follow different development trajectories. However, this seems unlikely given that participants were closely age matched across sites at all timepoints, reason why we believe a change in image quality across timepoints to be the most likely explanation for the observed residual differences.

Some limitations of this study should be acknowledged. The three included scanners were similar, but differed in gradient systems, generation of spectrometers, software versions and capabilities (e.g., multiband), allowing us to showcase the need for harmonization even when using hardware from a single vendor and almost identical acquisition parameters. On the other hand, however, this design did not allow us to evaluate the effectiveness of our method to harmonize multi-shell dMRI across different vendors. The effectiveness of the RISH method essentially depends on the maximum order of spherical harmonics that can be fitted to the data. The acquisition of at least 15 unique directions per-shell is needed to fit spherical harmonics of order 4, which is the minimum requirement to capture anisotropy details beyond the diffusion tensor. In this study, for example, we have discarded 10 out of 13 data shells originally acquired in the REPIMPACT study featuring less than 15 directions, although their inclusion would have likely been beneficial to improve the quality of fit. On the other hand, future studies should investigate which order of spherical harmonics is needed to effectively represent data acquired at small diffusion weighting, given that the angular resolution of dMRI data decreases with the diffusion weighting. Should harmonics of order 0 or 2 prove sufficient, this would be beneficial for the application of methods such as spectral intra-voxel incoherent motion (De Luca et al., 2018) and free-water (Pasternak et al., 2009) relying on the acquisition of a small number of directions at multiple diffusion weightings. The minimum number of directions to acquire per shell essentially depends on the desired application. In this work, spherical harmonics of order 4 (b = 1000s/mm^2) and 6 (b = 2500s/mm^2) have proven sufficient to preserve the quality of the DKI fit, but these values would have been likely suboptimal to further apply methods such as spherical deconvolution. For this reason, we suggest to acquire at least 45 unique directions per shell to fit spherical harmonics of at least order 8 (Tournier et al., 2007). Of note, DKI, NODDI or even some methods to quantity fiber orientations (Baete et al., 2019; Guo et al., 2020; Morez et al., 2021; Yeh et al., 2010) can also be applied to data acquired with non-shelled protocols, e.g., acquiring many diffusion weightings along few directions, similar to MAP-MRI (Özarslan et al., 2013) or CHARMED (Assaf and Basser, 2005), but these schemes are not compatible with the RISH method at this stage. An aspect which remains unclear is to which extent the RISH method can tolerate differences in acquisition protocols while still effectively harmonizing multi-shell data. In a recent study in patients with small vessel disease, we have shown that the RISH method could well harmonize DTI metrics derived from dMRI data collected in 5 centers with different MRI vendors, echo time, imaging resolution and diffusion weighting (de Brito Robalo et al., 2021). However, this is unlikely to be directly translatable to multi-shell protocols, given methods such as DKI and NODDI are likely more sensitive than DTI to differences in SNR, compartment weighting due to different echo times, and in how diffusion is effectively weighted (e.g., long weak gradients vs short strong gradients).

Despite the use of research dedicated MRI scanners, the data of one of the acquisition sites were characterized by the presence of non-negligible artefacts (S3). This represents, however, a real use case for the harmonization procedure, which overall performed remarkably well despite these limitations. Another aspect to be considered is the number of subjects needed to properly train the harmonization in order to average out individual trainings and only remove hardware related differences. Previous work has shown that at least 16 healthy controls – matched for age, sex and any other relevant covariate – should be used for this step. Given the nature of our study sample, we did not consider sex or education as additional covariates during harmonization training, but these and other relevant factors should be properly balanced in the training groups when studying broader populations to ensure no biases are introduced in subsequent analyses. Furthermore, we used 15 datasets including multiple acquisitions of some subjects to train the harmonization of S1, but only 11 datasets from 4 subjects to train the harmonization of S3. While this approach is arguably suboptimal, our results suggest that including data from multiple time points in the training might be a reasonable compromise to achieve proper harmonization when a limited number of unique subjects is available. Such scenarios may also arise, for example, when harmonizing data of retrospective studies where there is only a partial overlap in demographic characteristics of controls such as age and sex across sites. Of note, this approach seemed particularly suited to the harmonization of DKI metrics, whereas residual differences in NODDI metrics post-harmonization of S3 but not of S2 suggest the need for a larger number of unique subjects during training to effectively harmonize NODDI metrics. Finally, in this study we employed a linear least-squares fit for the kurtosis tensor, which is sensitive to the presence of outliers and artefacts in the data. The use of more advanced fitting methods based on weighted least squares (Veraart et al., 2013) or outlier rejection (Tax et al., 2015; Zhang et al., 2019) might further improve the comparability of inter-site data.

In conclusion, in this work we have shown that harmonizing individual diffusion weighted shells effectively attenuates inter-site offsets in DKI and NODDI metrics, yet does not alter their changes over longitudinal acquisitions. Altogether, our results suggest that the RISH harmonization method is suitable to pool multi-center dMRI data to increase sample sizes for multi-shell analyses, provided the data were acquired with comparable acquisition protocols.

Availability statement

The anonymized data used in this study will be made available to qualified researchers upon reasonable request to the authors.

Credit statement

Alberto De Luca: Conceptualization, Data curation, Methodology, Formal analysis, Investigation, Visualization, Project administration, Writing – original draft, writing – review & editing.

Suheyla Cetin Karayumak: Conceptualization, writing – review & editing, software development, Methodology.

Alexander Leemans: Conceptualization, Funding acquisition, writing – review & editing, Writing – original draft.

Yogesh Rathi: Conceptualization, Funding acquisition, writing – review & editing, software development, Methodology.

Stephan Swinnen: Conceptualization, Funding acquisition, writing – review & editing.

Jolien Gooijers: Data collection, Data curation, Conceptualization, writing – review & editing, Writing – original draft.

Amanda Clauwaert: Data collection, Data curation, Conceptualization, writing – review & editing.

Roald Bahr: Conceptualization, Funding acquisition, writing – review & editing.

Stian Bahr Sandmo: Data collection, Data curation, Conceptualization, writing – review & editing.

Nir Sochen: Conceptualization, Funding acquisition, writing – review & editing.

David Kaufmann: Data collection, Data curation, Conceptualization, writing – review & editing.

Marc Muehlmann: Data collection, Data curation, Conceptualization, writing – review & editing.

Geert-Jan Biessels: Conceptualization, Funding acquisition, writing – review & editing.

Inga Koerte: Conceptualization, Funding acquisition, writing – review & editing, Project administration, Writing – original draft.

Ofer Pasternak: Conceptualization, Funding acquisition, writing – review & editing, Project administration, Methodology, Writing – original draft.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119439.

References

Allen, N., Sudlow, C., Downey, P., Peakman, T., Danesh, J., Elliott, P., Gallacher, J., Green, J., Matthews, P., Pell, J., Sprosen, T., Collins, R., 2012. UK Biobank: current status and what it means for epidemiology. Heal. Policy Technol. 1, 123–126. doi:10.1016/j.hlpt.2012.07.003.

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20, 870–888. doi:10.1016/S1053-8119(03)00336-7.

Andersson, J.L.R., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063–1078. doi:10.1016/j.neuroimage.2015.10.019.

Arab, A., Wojna-Pelczar, A., Khairnar, A., Szabó, N., Ruda-Kucerova, J., 2018. Principles of diffusion kurtosis imaging and its role in early diagnosis of neurodegenerative disorders. Brain Res. Bull. 139, 91–98. doi:10.1016/j.brainresbull.2018.01.015.

Arfanakis, K., Haughton, V.M., Carew, J.D., Rogers, B.P., Dempsey, R.J., Meyerand, M.E., 2002. Diffusion tensor MR imaging in diffuse axonal injury. Am. J. Neuroradiol. 23, 794–802.

Assaf, Y., Basser, P.J., 2005. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. Neuroimage 27, 48–58. doi:10.1016/j.neuroimage.2005.03.042.

Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C., 2014. The Insight ToolKit image registration framework. Front. Neuroinform. 8, 44. doi:10.3389/fninf.2014.00044.

Baete, S.H., Cloos, M.A., Lin, Y.C., Placantonakis, D.G., Shepherd, T., Boada, F.E., 2019. Fingerprinting Orientation Distribution Functions in diffusion MRI detects smaller crossing angles. Neuroimage 198, 231–241. doi:10.1016/j.neuroimage.2019.05.024.

Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. B 111, 209–219. doi:10.1016/j.jmr.2011.09.022.

Calabrese, M., Castellaro, M., Bertoldo, A., De Luca, A., Pizzini, F.B., Ricciardi, G.K., Pitteri, M., Zimatore, S., Magliozzi, R., Benedetti, M.D., Manganotti, P., Montemezzi, S., Reynolds, R., Gajofatto, A., Monaco, S., 2017. Epilepsy in multiple sclerosis: the role of temporal lobe damage. Mult. Scler. 23, 473–482. doi:10.1177/1352458516651502.

Cetin Karayumak, S., Bouix, S., Ning, L., James, A., Crow, T., Shenton, M., Kubicki, M., Rathi, Y., 2019. Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. Neuroimage 184, 180–200. doi:10.1016/j.neuroimage.2018.08.073.

Chung, A.W., Seunarine, K.K., Clark, C.A., 2016. NODDI reproducibility and variability with magnetic field strength: a comparison between 1.5 T and 3 T. Hum. Brain Mapp 37, 4550–4565. doi:10.1002/hbm.23328.

Cole, J.H., 2020. Multi-modality neuroimaging brain-age in UK Biobank: relationship to biomedical, lifestyle and cognitive factors. Neurobiol. Aging 92, 34–42. doi:10.1016/j.neurobiolaging.2020.03.014.

Colon-Perez, L.M., Ibanez, K.R., Suarez, M., Torroella, K., Acuna, K., Ofori, E., Levites, Y., Vaillancourt, D.E., Golde, T.E., Chakrabarty, P., Febo, M., 2019. Neurite orientation dispersion and density imaging reveals white matter and hippocampal microstructure changes produced by Interleukin-6 in the TgCRND8 mouse model of amyloidosis. Neuroimage 202, 116138. doi:10.1016/j.neuroimage.2019.116138.

Das, S.K., Wang, J.L., Bing, L., Bhetuwal, A., Yang, H.F., 2017. Regional Values of Diffusional Kurtosis Estimates in the Healthy Brain during Normal Aging. Clin. Neuroradiol. 27, 283–298. doi:10.1007/s00062-015-0490-z.

de Brito Robalo, B.M., Biessels, G.J., Chen, C., Dewenter, A., Duering, M., Hilal, S., Koek, H.L., Kopczak, A., Yin Ka Lam, B., Leemans, A., Mok, V., Onkenhout, L.P., van den Brink, H., de Luca, A., 2021. Diffusion MRI harmonization enables joint-analysis of multicentre data of patients with cerebral small vessel disease. NeuroImage Clin 32, 102886. doi:10.1016/j.nicl.2021.102886.

De Luca, A., et al., 2021. On the generalizability of diffusion MRI signal representations across acquisition parameters, sequences and tissue types: Chronicles of the MEMENTO challenge. Neuroimage 240. doi:10.1016/j.neuroimage.2021.118367.

De Luca, A., Bertoldo, A., Froeling, M., 2017. Effects of perfusion on DTI and DKI estimates in the skeletal muscle. Magn. Reson. Med. 78, 233–246. doi:10.1002/mrm.26373.

de Luca, A., Biessels, G.J., 2021. Towards multicentre diffusion MRI studies in cerebral small vessel disease. J. Neurol. Neurosurg. Psychiatry 0. doi:10.1136/jnnp-2021-326993, jnnp-2021-326993.

De Luca, A., Guo, F., Froeling, M., Leemans, A., 2020. Spherical deconvolution with tissue-specific response functions and multi-shell diffusion MRI to estimate multiple fiber orientation distributions (mFODs). Neuroimage 222, 117206. doi:10.1016/j.neuroimage.2020.117206.

De Luca, A., Leemans, A., Bertoldo, A., Arrigoni, F., Froeling, M., 2018. A robust deconvolution method to disentangle multiple water pools in diffusion MRI. NMR Biomed 1–17. doi:10.1002/nbm.3965.

Dewey, B.E., Zhao, C., Reinhold, J.C., Carass, A., Fitzgerald, K.C., Sotirchos, E.S., Saidha, S., Oh, J., Pham, D.L., Calabresi, P.A., van Zijl, P.C.M., Prince, J.L., 2019. DeepHarmony: a deep learning approach to contrast harmonization across scanner changes. Magn. Reson. Imaging 64, 160–170. doi:10.1016/j.mri.2019.05.041.

Falangola, M.F., Jensen, J.H., Babb, J.S., Hu, C., Castellanos, F.X., Di Martino, A., Ferris, S.H., Helpern, J.A., 2008. Age-related non-Gaussian diffusion patterns in the prefrontal brain. J. Magn. Reson. Imaging 28, 1345–1350. doi:10.1002/jmri.21604.

Fieremans, E., Jensen, J.H., Helpern, J.a, 2011. White matter characterization with diffusional kurtosis imaging. Neuroimage 58, 177–188. doi:10.1016/j.neuroimage.2011.06.006.

Fortin, J.P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 167, 104–120. doi:10.1016/j.neuroimage.2017.11.024.

Girault, J.B., Cornea, E., Goldman, B.D., Knickmeyer, R.C., Styner, M., Gilmore, J.H., 2019. White matter microstructural development and cognitive ability in the first 2 years of life. Hum. Brain Mapp 40, 1195–1210. doi:10.1002/hbm.24439.

Guo, F., Leemans, A., Viergever, M.A., Dell'Acqua, F., De Luca, A., 2020. Generalized Richardson-Lucy (GRL) for analyzing multi-shell diffusion MRI data. Neuroimage 218, 116948. doi:10.1016/j.neuroimage.2020.116948.

Gyebnár, G., Szabó, Á., Sirály, E., Fodor, Z., Sákovics, A., Salacz, P., Hidasi, Z., Csibri, É., Rudas, G., Kozák, L.R., Csukly, G., 2018. What can DTI tell about early cognitive impairment? – Differentiation between MCI subtypes and healthy controls by diffusion tensor imaging. Psychiatry Res. - Neuroimaging 272, 46–57. doi:10.1016/j.pscychresns.2017.10.007.

Harms, R.L., et al., 2017. Robust and fast nonlinear optimization of diffusion MRI microstructure models. Neuroimage 155. doi:10.1016/j.neuroimage.2017.04.064.

Jensen, J.H., Helpern, J.a, 2010. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR Biomed 23, 698–710. doi:10.1002/nbm.1518.

Jensen, J.H., Helpern, J.a, Ramani, A., Lu, H., Kaczynski, K., 2005. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn. Reson. Med. 53, 1432–1440. doi:10.1002/mrm.20508.

Koerte, I.K., Bahr, R., Filipcik, P., Gooijers, J., Leemans, A., Lin, A.P., Tripodis, Y., Shenton, M.E., Sochen, N., Swinnen, S.P., Pasternak, O., 2021. REPIMPACT - a prospective longitudinal multisite study on the effects of repetitive head impacts in youth soccer. Brain Imaging Behav doi:10.1007/s11682-021-00484-x.

Konieczny, M.J., Dewenter, A., Ter Telgte, A., Gesierich, B., Wiegertjes, K., Finsterwalder, S., Kopczak, A., Hübner, M., Malik, R., Tuladhar, A.M., Marques, J.P., Norris, D.G., Koch, A., Dietrich, O., Ewers, M., Schmidt, R., de Leeuw, F.-.E., Duering, M., 2021. Multi-shell Diffusion MRI Models for White Matter Characterization in Cerebral Small Vessel Disease. Neurology 96, e698–e708. doi:10.1212/WNL.0000000000011213.

Lampinen, B., Szczepankiewicz, F., Novén, M., van Westen, D., Hansson, O., Englund, E., Mårtensson, J., Westin, C.F., Nilsson, M., 2019. Searching for the neurite density with diffusion MRI: challenges for biophysical modeling. Hum. Brain Mapp. 40, 2529–2545. doi:10.1002/hbm.24542.

Lebel, C., Walker, L., Leemans, A., Phillips, L., Beaulieu, C., 2008. Microstructural maturation of the human brain from childhood to adulthood. Neuroimage 40, 1044–1055. doi:10.1016/j.neuroimage.2007.12.053.

Leemans, A., Jeurissen, B., Sijbers, J., Jones, D.K., 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In: 17th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii, USA, p. 3537.

Lehmann, N., Aye, N., Kaufmann, J., Heinze, H.J., Düzel, E., Ziegler, G., Taubert, M., 2021. Longitudinal Reproducibility of Neurite Orientation Dispersion and Density Imaging (NODDI) Derived Metrics in the White Matter. Neuroscience 457, 165–185. doi:10.1016/j.neuroscience.2021.01.005.

Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Grant, G., Marx, C.E., Morey, R.A., Flashman, L.A., George, M.S., McAllister, T.W., Andaluz, N., Shutter, L., Coimbra, R., Zafonte, R.D., Coleman, M.J., Kubicki, M., Westin, C.F., Stein, M.B., Shenton, M.E., Rathi, Y., 2016. Inter-site and inter-scanner diffusion MRI data harmonization. Neuroimage 135, 311–323. doi:10.1016/j.neuroimage.2016.04.041.

Morez, J., Sijbers, J., Vanhevel, F., Jeurissen, B., 2021. Constrained spherical deconvolution of nonspherically sampled diffusion MRI data. Hum. Brain Mapp. 42, 521–538. doi:10.1002/hbm.25241.

Ning, L., Bonet-Carne, E., Grussu, F., Sepehrband, F., Kaden, E., Veraart, J., Blumberg, S.B., Khoo, C.S., Palombo, M., Kokkinos, I., Alexander, D.C., Coll-Font, J., Scherrer, B., Warfield, S.K., Karayumak, S.C., Rathi, Y., Koppers, S., Weninger, L., Ebert, J., Merhof, D., Moyer, D., Pietsch, M., Christiaens, D., Gomes Teixeira, R.A., Tournier, J.D., Schilling, K.G., Huo, Y., Nath, V., Hansen, C., Blaber, J., Landman, B.A., Zhylka, A., Pluim, J.P.W., Parker, G., Rudrapatna, U., Evans, J., Charron, C., Jones, D.K., Tax, C.M.W., 2020. Cross-scanner and cross-protocol multi-shell diffusion MRI data harmonization: algorithms and results. Neuroimage 221. doi:10.1016/j.neuroimage.2020.117128.

Ning, L., Setsompop, K., Michailovich, O., Makris, N., Westin, C.F., Rathi, Y., 2015. A compressed-sensing approach for super-resolution reconstruction of diffusion MRI. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9123, 57–68. doi:10.1007/978-3-319-19992-4_5.

Özarslan, E., Koay, C.G., Shepherd, T.M., Komlosh, M.E., İrfanoğlu, M.O., Pierpaoli, C., Basser, P.J., 2013. Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. Neuroimage 78, 16–32. doi:10.1016/j.neuroimage.2013.04.016.

Parvathaneni, P., Rogers, B.P., Huo, Y., Schilling, K.G., Hainline, A.E., Anderson, A.W., Woodward, N.D., Landman, B.A., 2017. Gray matter surface based spatial statis-

tics (GS-BSS) in diffusion microstructure. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10433 LNCS 638–646. doi:10.1007/978-3-319-66182-7_73.

Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y., 2009. Free water elimination and mapping from diffusion MRI. Magn. Reson. Med. 62, 717–730. doi:10.1002/mrm.22055.

Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I.M., Satterthwaite, T.D., Fan, Y., Launer, L.J., Masters, C.L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S.C., Fripp, J., Koutsouleris, N., Wolf, D.H., Gur, Raquel, Gur, Ruben, Morris, J., Albert, M.S., Grabe, H.J., Resnick, S.M., Bryan, R.N., Wolk, D.A., Shinohara, R.T., Shou, H., Davatzikos, C., 2020. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage 208. doi:10.1016/j.neuroimage.2019.116450.

Setsompop, K., Fan, Q., Stockmann, J., Bilgic, B., Huang, S., Cauley, S.F., Nummenmaa, A., Wang, F., Rathi, Y., Witzel, T., Wald, L.L., 2018. High-Resolution In Vivo Diffusion Imaging of the Human Brain With Generalized Slice Dithered Enhanced Resolution: simultaneous Multislice (gSlider-SMS) 151, 141–151. https://doi.org/10.1002/mrm.26653

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155. doi:10.1002/hbm.10062.

Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487–1505. doi:10.1016/j.neuroimage.2006.02.024.

Smith, S.M., Nichols, T.E., 2018. Statistical Challenges in “Big Data” Human Neuroimaging. Neuron 97, 263–268. doi:10.1016/j.neuron.2017.12.018.

St-Jean, S., Viergever, M.A., Leemans, A., 2020. Harmonization of diffusion MRI data sets with adaptive dictionary learning. Hum. Brain Mapp. 41, 4478–4499. doi:10.1002/hbm.25117.

Tamnes, C.K., Roalf, D.R., Goddings, A.L., Lebel, C., 2018. Diffusion MRI of white matter microstructure development in childhood and adolescence: methods, challenges and progress. Dev. Cogn. Neurosci. 33, 161–175. doi:10.1016/j.dcn.2017.12.002.

Tax, C.M., Grussu, F., Kaden, E., Ning, L., Rudrapatna, U., John Evans, C., St-Jean, S., Leemans, A., Koppers, S., Merhof, D., Ghosh, A., Tanno, R., Alexander, D.C., Zappalà, S., Charron, C., Kusmia, S., Linden, D.E., Jones, D.K., Veraart, J., 2019. Cross-scanner and cross-protocol diffusion MRI data harmonisation: a benchmark database and evaluation of algorithms. Neuroimage 195, 285–299. doi:10.1016/j.neuroimage.2019.01.077.

Tax, C.M.W., Otte, W.M., Viergever, M.A., Dijkhuizen, R.M., Leemans, A., 2015. REKINDLE: robust extraction of kurtosis INDices with linear estimation. Magn. Reson. Med. 73, 794–808. doi:10.1002/mrm.25165.

Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35, 1459–1472. doi:10.1016/j.neuroimage.2007.02.016.

Turner, B.O., Paul, E.J., Miller, M.B., Barbey, A.K., 2018. Small sample sizes reduce the replicability of task-based fMRI studies. Commun. Biol. 1. doi:10.1038/s42003-018-0073-z.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K.WU-Minn HCP Consortium, 2013. The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041.

Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B., 2013. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. Neuroimage 81, 335–346. doi:10.1016/j.neuroimage.2013.05.028.

Vos, S.B., Tax, C.M.W., Luijten, P.R., Ourselin, S., Leemans, A., Froeling, M., 2016. The importance of correcting for signal drift in diffusion MRI. Magn. Reson. Med. 22, 4460. doi:10.1002/mrm.26124.

Wen, Q., Mustafi, S.M., Li, J., Risacher, S.L., Tallman, E., Brown, S.A., West, J.D., Harezlak, J., Farlow, M.R., Unverzagt, F.W., Gao, S., Apostolova, L.G., Saykin, A.J., Wu, Y.C., 2019. White matter alterations in early-stage Alzheimer’s disease: a tract-specific study. Alzheimer’s Dement. Diagnosis, Assess. Dis. Monit. 11, 576–587. doi:10.1016/j.dadm.2019.06.003.

Yeh, F.-C., Wedeen, V.J., Tseng, W.-Y.I., 2010. Generalized q-sampling imaging. IEEE Trans. Med. Imaging 29, 1626–1635. doi:10.1109/TMI.2010.2045126.

Zhang, F., Ning, L., O’Donnell, L.J., Pasternak, O., 2019. MK-curve - Characterizing the relation between mean kurtosis and alterations in the diffusion MRI signal. Neuroimage 196, 68–80. doi:10.1016/j.neuroimage.2019.04.015.