



Diffusion MRI harmonization enables joint-analysis of multicentre data of patients with cerebral small vessel disease

Bruno M. de Brito Robalo^{a,b,*}, Geert Jan Biessels^a, Christopher Chen^c, Anna Dewenter^d, Marco Duering^{d,e}, Saima Hilal^c, Huiberdina L. Koek^f, Anna Kopczak^d, Bonnie Yin Ka Lam^g, Alexander Leemans^b, Vincent Mok^g, Laurien P. Onkenhout^a, Hilde van den Brink^a, Alberto de Luca^{a,b}

^a Department of Neurology and Neurosurgery, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands

^b Image Sciences Institute, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

^c Memory, Aging and Cognition Center, Department of Pharmacology, National University of Singapore, Singapore

^d Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Germany

^e Medical Image Analysis Center (MIAC AG) and qbig, Department of Biomedical Engineering, University of Basel, Basel, Switzerland

^f Department of Geriatric Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

^g Division of Neurology, Department of Medicine and Therapeutics, Gerald Choa Neuroscience Centre, Faculty of Medicine, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:

Diffusion MRI
Harmonization
Cerebral small vessel disease
Multicentre
White matter hyperintensities

ABSTRACT

Objectives: Acquisition-related differences in diffusion magnetic resonance imaging (dMRI) hamper pooling of multicentre data to achieve large sample sizes. A promising solution is to harmonize the raw diffusion signal using rotation invariant spherical harmonic (RISH) features, but this has not been tested in elderly subjects. Here we aimed to establish if RISH harmonization effectively removes acquisition-related differences in multicentre dMRI of elderly subjects with cerebral small vessel disease (SVD), while preserving sensitivity to disease effects. **Methods:** Five cohorts of patients with SVD (N = 397) and elderly controls (N = 175) with 3 Tesla MRI on different systems were included. First, to establish effectiveness of harmonization, the RISH method was trained with data of 13 to 15 age and sex-matched controls from each site. Fractional anisotropy (FA) and mean diffusivity (MD) were compared in matched controls between sites using tract-based spatial statistics (TBSS) and voxel-wise analysis, before and after harmonization. Second, to assess sensitivity to disease effects, we examined whether the contrast (effect sizes of FA, MD and peak width of skeletonized MD - PSMD) between patients and controls within each site remained unaffected by harmonization. Finally, we evaluated the association between white matter hyperintensity (WMH) burden, FA, MD and PSMD using linear regression analyses both within individual cohorts as well as with pooled scans from multiple sites, before and after harmonization. **Results:** Before harmonization, significant differences in FA and MD were observed between matched controls of different sites ($p < 0.05$). After harmonization these site-differences were removed. Within each site, RISH harmonization did not alter the effect sizes of FA, MD and PSMD between patients and controls (relative change in Cohen's $d = 4\%$) nor the strength of association with WMH volume (relative change in $R^2 = 2.8\%$). After harmonization, patient data of all sites could be aggregated in a single analysis to infer the association between WMH volume and FA ($R^2 = 0.62$), MD ($R^2 = 0.64$), and PSMD ($R^2 = 0.60$). **Conclusions:** We showed that RISH harmonization effectively removes acquisition-related differences in dMRI of elderly subjects while preserving sensitivity to SVD-related effects. This study provides proof of concept for future multicentre SVD studies with pooled datasets.

* Corresponding author at: Department of Neurology and Neurosurgery, UMC Utrecht Brain Center, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands.

E-mail addresses: B.M.deBritoRobalo-2@umcutrecht.nl (B.M. de Brito Robalo), G.J.Biessels@umcutrecht.nl (G.J. Biessels), phccclh@nus.edu.sg (C. Chen), Anna.Dewenter@med.uni-muenchen.de (A. Dewenter), marco.duering@miac.ch (M. Duering), saimahilal@nus.edu.sg (S. Hilal), H.L.Koek@umcutrecht.nl (H.L. Koek), Anna.Kopczak@med.uni-muenchen.de (A. Kopczak), bonnieyklam@cuhk.edu.hk (B. Yin Ka Lam), A.Leemans@umcutrecht.nl (A. Leemans), vctmok@cuhk.edu.hk (V. Mok), L.P.Onkenhout@umcutrecht.nl (L.P. Onkenhout), H.vandenBrink-2@umcutrecht.nl (H. van den Brink), A.deLuca-2@umcutrecht.nl (A. de Luca).

¹ ORCID: 0000-0001-9957-5039.

<https://doi.org/10.1016/j.nicl.2021.102886>

Received 21 October 2021; Accepted 16 November 2021

Available online 18 November 2021

2213-1582/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Combining data from multicentre studies is becoming increasingly important in neuroimaging, with the aim to increase statistical power and provide outcomes that are more generalizable than those obtained at single-centre level (Lorca-Puls et al., 2018; Van Horn & Toga, 2014). However, joint analysis of multicentre magnetic resonance imaging (MRI) can be challenging if inter-site variability due to acquisition-related inconsistencies is not taken into account (Vollmar et al., 2010).

Inter-site variability is particularly problematic in diffusion MRI (dMRI) and can be caused by a range of factors, including scanner hardware (e.g., scanner manufacturer, magnetic field strength, gradient strength, field inhomogeneities), software, acquisition parameters (e.g., voxel size, number of gradient directions, echo time) (Helmer et al., 2016). All these factors may affect the measured diffusion signal intensity and metrics derived from the data. In prospective multicentre studies, this variability can be controlled using standardized acquisitions and scanners from the same manufacturer (Konieczny et al., 2020). However, when retrospectively combining data from different cohorts, differences in acquisition can be substantial. Even with phantoms, dMRI metrics have shown more than 7% variability across sites (Palacios et al., 2017; Teipel et al., 2011; Timmermans et al., 2019). In the human brain, this variability is even more pronounced and non-uniform across tissues with acquisition-related differences reaching the same order of magnitude as case-control differences (e.g., traumatic brain injury vs. controls, Kumar et al., 2009). In such scenarios, if multicentre data were naively pooled into a single analysis, true biological effects would likely be masked by acquisition-related differences. It is therefore crucial that multicentre dMRI is harmonized prior to joint-analysis (Tax et al., 2019).

Two main kinds of retrospective dMRI harmonization techniques have been developed to date. The first category operates on each diffusion metric individually (e.g., fractional anisotropy – FA, and mean diffusivity – MD) by using statistical approaches such as *meta*-analysis, or by modelling the difference between sites with covariates during analysis (e.g., ComBat, Fortin et al., 2017). By contrast, the second category of harmonization operates directly on the raw dMRI data rather than on each diffusion metric (Karayumak et al., 2019; Koppers et al., 2019; Mirzaalian et al., 2015). This type of harmonization is more general since the raw diffusion signal is harmonized in a model-independent manner, theoretically allowing any type of subsequent analysis. In this study, we focus specifically on the second type of harmonization with the rotation invariant spherical harmonics (RISH) methods (Mirzaalian et al., 2015).

The core idea of the RISH method is to map the dMRI signal from a ‘target’ site to a ‘reference’ site, using groups of healthy subjects matched for factors such as age, sex, etc. This signal mapping is possible because the dMRI signal intensity can be represented in a spherical harmonics (SH) basis with a given number of parametrization coefficients (Tournier et al., 2004). From this representation, RISH features can be extracted and scaled to harmonize the dMRI signal between two sites (Karayumak et al., 2019). Applications of RISH harmonization have been presented using synthetic data and with data of healthy young subjects, with recent work showing that acquisition-related differences are removed while preserving age- and sex-related effects (Karayumak et al., 2019). Recently, this method has also been applied to harmonize a large dataset of patients with Schizophrenia and investigate changes across the lifespan (Cetin-Karayumak et al., 2020). However, the applicability of the RISH method to older individuals exhibiting brain atrophy or to patients with (diffuse) white matter lesions remain unclear.

In this work, we evaluate the RISH harmonization framework in the context of a retrospective multicentre analysis of individuals with brain lesions due to cerebral small vessel disease (SVD). SVD is a leading cause of cognitive impairment and dementia and it is often investigated with dMRI (Baykara et al., 2016; Lyoubi-Idrissi et al., 2017; Wiegertjes et al.,

2019). The patterns of diffusion change in SVD are well documented using the diffusion tensor model, with patients typically exhibiting widespread increase of MD, peak width of skeletonized MD (PSMD) and decrease of FA, often related to white matter hyperintensity (WMH) burden (Tuladhar et al., 2015; Van Norden et al., 2012). Hence, this patient population is well-suited to investigate the efficacy of harmonization methods. Using scans from five SVD cohorts acquired on different systems (Philips Healthcare, NL, and Siemens Healthineers, DE) and with different protocols, we aimed to establish if application of the RISH method removes acquisition-related differences in dMRI of elderly subjects, while preserving the sensitivity to disease effects in SVD. Finally, we show proof of concept of how multicentre harmonized data can be pooled to perform robust inference of the relation between WMH burden and dMRI metrics.

2. Methods

2.1. Datasets and inclusion criteria

For this retrospective analysis, we obtained scans from five cohorts including healthy elderly subjects and patients with SVD. These cohorts differed in study design and inclusion criteria (described below), comprising four samples with sporadic SVD and one sample with genetically defined SVD (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy, CADASIL). For the present study, we used a harmonized definition for patients and controls. Patients with sporadic SVD had symptomatic SVD defined as a) history of stroke, with a corresponding small subcortical infarct visible on MRI or b) cognitive complaints and presence of WMH burden on MRI (Fazekas score ≥ 2 , Fazekas et al., 1987). The presence of CADASIL was confirmed by molecular genetic testing of the *NOTCH3* gene or ultra-structural analysis of a skin biopsy (detection of pathognomonic granular osmiophilic material, Wollenweber et al., 2015). Patients were excluded if they had other major neurological or psychiatric conditions (e.g., multiple sclerosis, epilepsy, Parkinson’s disease). Controls had no history of stroke or cognitive complaints for which they sought medical advice, and their MRI did not show signs of lacunes or WMH with Fazekas score ≥ 2 . All subjects had a structural MRI (T1-weighted) and a dMRI scan. Characteristics of the study samples included in this study (397 patients and 175 controls) are provided in Tables 1. All studies included in this analysis were approved by the ethics committees of the respective institutions and all participants provided written informed consent.

2.1.1. Utrecht1

Patients ($n = 171$) were selected from the Parelstoer study memory clinic cohort (Aalten et al., 2014). Age-matched controls ($n = 53$) were recruited from a community-based cohort (Reijmer et al., 2013). All MRI scans from both cohorts were acquired on the same 3 Tesla Philips scanner (Achieva, Philips, Best, the Netherlands). T1-weighted scans for both cohorts were acquired with the following parameters: voxel size: $1 \times 1 \times 1 \text{ mm}^3$, echo time (TE): 4.5 ms and repetition time (TR): 7.9 ms. dMRI data were obtained with a voxel size: $2.5 \times 2.5 \times 2.5 \text{ mm}^3$, TR/TE 6638/73 ms, 45 diffusion gradients directions with a b-value of 1200 s/mm^2 , and $1b = 0 \text{ s/mm}^2$ averaged 3 times. Fluid-attenuated inversion recovery (FLAIR) images were obtained with TR/TE/inversion time (TI): 11000/125/2800 ms, voxel size: $1 \times 1 \times 3 \text{ mm}^3$.

2.1.2. Hong Kong

Patients ($n = 20$) and controls ($n = 20$) were selected from a community-based cohort, the Chinese University of Hong Kong–Risk Index for Subclinical brain lesions in Hong Kong (CU-RISK) (Lam et al., 2019). MRI scans were acquired on a 3 Tesla Philips scanner (Achieva, Philips, Best, the Netherlands). T1-weighted images were obtained with TR/TE: 7.49/3.46 ms, voxel size: $0.60 \times 1.04 \times 1.04 \text{ mm}^3$ and dMRI had a TR/TE: 8944/60 ms, voxel size: $1 \times 1 \times 2 \text{ mm}^3$; 32 diffusion gradient

directions with b-value 1000 s/mm² and 1b = 0 s/mm². FLAIR images were acquired with TR/TE/TI: 8000/328.6/2400 ms, voxel size: 0.55 × 0.44 × 0.44 mm³.

2.1.3. Munich

Patients (n = 72) with CADASIL and controls (n = 34) were selected from the prospective VASCAMY (Vascular and Amyloid Predictors of Neurodegeneration and Cognitive Decline in Nondemented Subjects) study (Baykara et al., 2016). All MRI scans were acquired on a 3 Tesla Magnetom Verio scanner (Siemens Healthineers, Erlangen, Germany). T1-weighted scans were obtained using TR/TE: 2500/4.73 ms, voxel size: 1 × 1 × 1 mm³ and dMRI were acquired with a voxel size: 2 × 2 × 2 mm³, TR/TE: 12700/81 ms, 30 diffusion gradient directions with a b-value of 1000 s/mm², and 1b = 0 s/mm². FLAIR images were obtained with TR/TE/TI: 5000/395/1800 ms, voxel size: 1 × 1 × 1 mm³.

2.1.4. Utrecht2

A second dataset from the UMC Utrecht consisted of patients (n = 34) and controls (n = 18) from an ongoing prospective observational cohort study Zoom@SVDs (van den Brink et al., 2021). MRI scans were acquired using the same scanner system and acquisition parameters as the Utrecht1 dataset. However, since multiple scanner software and hardware (coil) updates occurred between the two studies, scans from the Zoom@SVDs study are treated as a separate site.

2.1.5. Singapore

Patients (n = 100) and controls (n = 50) were selected from a community-based cohort, the Epidemiology of Dementia In Singapore (EDIS) study (Hilal et al., 2013). All MRI scans were performed on a 3 Tesla Siemens Magnetom Trio Tim scanner (Siemens Healthineers, Erlangen, Germany). T1-weighted scans were obtained with TR/TE: 2300/1.9 ms, voxel size: 1 × 1 × 1 mm³ and dMRI were acquired with a TR/TE: 6800/85 ms, voxel size: 3.1 × 3.1 × 3 mm³; 61 diffusion gradient directions with b-value 1150 s/mm² and 7b = 0 s/mm². FLAIR images were obtained with TR/TE/inversion time (TI): 9000/82/2500 ms, voxel size: 1 × 1 × 3 mm³.

2.2. MRI data pre-processing

All datasets were pre-processed using ExploreDTI version 4.8.6 (Leemans et al., 2009) and the Functional Magnetic Resonance Imaging of the Brain (FMRIB) software library (FSL, v6.0.1). Images were corrected for signal drift (Vos et al., 2016), eddy currents, subject motion with rotation of the B-matrix (Leemans and Jones, 2009), and susceptibility induced distortions (Veraart et al., 2013). dMRI data were nonlinearly registered to the T1 and resampled to an isotropic resolution of 2 × 2 × 2 mm³ and brain masks were generated using brain extraction (BET) tool from FSL. All images were visually inspected to exclude the presence of major artifacts and misregistration. Since the dMRI data were acquired with different b-values in the different cohorts, we adjusted for differences in b-values as part of the harmonization pipeline. We estimated the signal of all dMRI data to a common b-value (b = 1000 s/mm²) using a linear scaling of the signal decay (S/S0) in the logarithmic domain (Jensen et al., 2005; Steven et al., 2014). This signal decay has been validated in phantoms and healthy controls and shown to be robust for datasets with closely spaced b-values centred around b = 1000 s/mm² where the diffusion signal is not heavily weighted towards non-Gaussian effects (Magin et al., 2019). This approach was also utilized by (Karayumak et al., 2019) for the RISH harmonization method. In our case, b-value scaling was applied to Utrecht1, Utrecht2 and Singapore, which had original b-values of 1200 s/mm² and 1150 s/mm².

WMH volumes were segmented from the FLAIR images using an automated pipeline (*coroflo*) and registered the MNI152 template (Kuijff et al., 2019). All volumes were normalized to the percentage of intracranial volume (ICV) of the MNI brain.

2.3. Harmonization with RISH features

Harmonization of dMRI with rotation invariant spherical harmonics (RISH) features was first proposed by Mirzaalian et al. 2015, with recent improvements allowing harmonization of scans with different acquisition parameters (Karayumak et al., 2019), which is the case in our study. This type of harmonization is based on the fact that dMRI signal along

Table 1
Demographics and imaging parameters of the study samples.

	Utrecht 1		Hong Kong		Munich		Utrecht 2		Singapore	
	Controls (N = 53)	Patients (N = 171)	Controls (N = 20)	Patients (N = 20)	Controls (N = 34)	Patients (N = 72)	Controls (N = 18)	Patients (N = 34)	Controls (N = 50)	Patients (N = 100)
Demographics										
Age, years	71.0 ± 4.8	74.9 ± 9.0	69.2 ± 3.4	74.1 ± 3.3	71.1 ± 4.4	53.4 ± 6.3	62.4 ± 6.9	66.9 ± 8.9	66.6 ± 4.8	72.6 ± 6.9
Male sex (%)	31 (58)	99 (59)	10 (50)	10 (50)	17 (49)	23 (32)	10 (55)	22 (65)	31 (52)	35 (35)
Cognitive testing										
MMSE	28 [28–30]	26 [24, 28]	–	–	30 [29, 30]	30 [27,30]	29 [28,30]	29 [27,30]	28 [27,29]	22 [18,25]
MoCA	–	–	25 [23, 27]	19 [14, 20]	–	–	–	–	27 [26,28]	11 [18,21]
MRI markers										
WMH volume (% ICV)	0.3 [0.1, 0.6]	1.2 [0.5, 2.7]	0.1 [0.05, 0.1]	0.6 [0.3, 1.1]	0.2 [0.1 0.7]	6.2 [3.5, 10]	0.03 [0.02, 0.08]	0.7 [0.3, 1.3]	0.04 [0.02, 0.09]	0.8 [0.08, 1.16]
WMH (Fazekas)	0 [0, 1]	2 [1, 2]	0.5 [0, 1]	2 [2, 3]	0 [0, 1]	3 [2, 3]	1 [1,1]	2 [2,3]	1 [1,1]	2 [1,3]
Lacunae (present)	0 (0)	69 (36)	0(0)	–	0 (0)	49 (6)	0(0)	17 (50)	0 (0)	37 (37)
Imaging parameters										
Scanner	3 T Philips Achieva		3 T Philips Achieva		3 T Siemens Verio		3 T Philips Achieva		3 T Siemens Magnetom Trio, Tim	
Software version	R3.1		MR Release 5.1		syngo MR B19		MR Release 5.6.0		syngo MR B19	
Voxel size (mm ²)	2.5 × 2.5 × 2.5		1 × 1 × 2		2 × 2 × 2		2.5 × 2.5 × 2.5		3.1 × 3.1 × 3.0	
b-value (s/mm ²)	1200		1000		1000		1200		1150	
# of directions	45		32		30		45		61	

Data presented as mean ± SD, number (percentages) or median [interquartile range]; SVD = small vessel disease; CADASIL = cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy; MMSE = Mini-Mental State Examination; MoCA = Montreal Cognitive Assessment; WMH = white matter hyperintensity; ICV = intracranial volume.

unique gradient directions can be represented with a basis of spherical harmonics (SH). From this representation, RISH features that describe different aspects of the signal can be calculated. RISH features can be viewed as the total energy at a specific angular frequency (order) in the SH space. The core assumption of this method is that two groups of healthy subjects matched for age, sex, lesion burden, etc., are expected to have similar diffusion profiles on a group level and thus none of the RISH features should be statistically different between sites. Under this assumption, eventual group differences observed in diffusion measurements such as FA and MD are attributed to scanner-related inconsistencies, as previously shown in healthy controls (Ning et al., 2020). To ensure that the average of RISH features captures site properties on a group level and not characteristics of individuals, a minimum number of training controls (15–20) is required from each site (Karayumak et al., 2019). Subsequently, a scaling is determined between the average of RISH features such that scanner-related differences are removed between sites. This mapping is linear in the SH domain, but non-linear in the original diffusion signal domain. We provide a detailed description of RISH harmonization in the [supplementary information](#), and further theory can also be found in the original method papers (Karayumak et al., 2019; Mirzaalian et al., 2015)

In short, RISH harmonization pipeline consists of two parts: 1) learning inter-site differences in the form of scale maps between RISH features of the reference and target site (Fig. 1, part 1); and 2) applying the learned scale maps to harmonize all dMRI datasets of the target site (Fig. 1, part 2). The learning part is performed using age- and sex-matched controls as training data. From the dMRI signal, the RISH features are calculated and registered to a common spatial template generated from training subjects with ANTs (Avants et al., 2010). In the template space, the expected values of RISH features are defined as the sample mean over the number of training subjects and voxel-wise scale maps of RISH features are estimated between the target and reference site. Next, in the application part, the scale maps are warped to the subject space and used to harmonize the SH coefficients of the target site. Finally, the harmonized diffusion signal can be reconstructed.

2.4. Experimental design and analysis

2.4.1. Effectiveness of RISH harmonization

Here we assess if acquisition-related differences in diffusion metrics between sites can be removed by RISH harmonization. The first step was to select Training Controls from every cohort that were as similar as possible between sites to minimize sources of variability other than scanner. The Utrecht1 cohort was used as a reference site because the age range of the controls allowed matching with all other sites. This was done on a site-by-site basis, generating four sets of Training Controls with participants from every site matched for age and sex to participants from Utrecht1 (demographics in supplementary Table S1). Tract-based-spatial-statistics (TBSS, Smith et al., 2006) and voxel-based analysis were used to compare the FA and MD between Training Controls of the reference and target sites, before and after harmonization. For the TBSS pipeline, FA and MD were estimated using the diffusion tensor model (*dfit* from FSL). Next, FA maps were aligned to the MNI152 template and a white matter skeleton representing the centers of major bundles was computed. Subsequently, FA and MD of the skeleton were compared between reference and target sites in a voxel-wise fashion using t-tests with threshold-free cluster enhancement (5000 permutations). This comparison was also extended to the whole brain to ensure that acquisition-related differences in grey matter regions and other structures are also removed.

We also evaluated the generalizability of effectiveness of harmonization beyond the Training Controls. This was done by creating a group of Validation Controls with data from Utrecht1 ($n = 15$), Munich ($n = 15$) and Singapore ($n = 15$), since those sites had a sufficient number of controls outside the Training Controls to generate separate sets of matched groups (demographics in supplementary Table S2). Similar to the analysis with training controls, TBSS and voxel-based analysis were used to compare validation controls between each target site and the reference, before and after harmonization. Furthermore, one-way ANOVA was performed to compare the average FA and MD of the TBSS skeleton across these three sites.

2.4.2. Sensitivity to disease effects in SVD

For this objective we included all patients from all sites

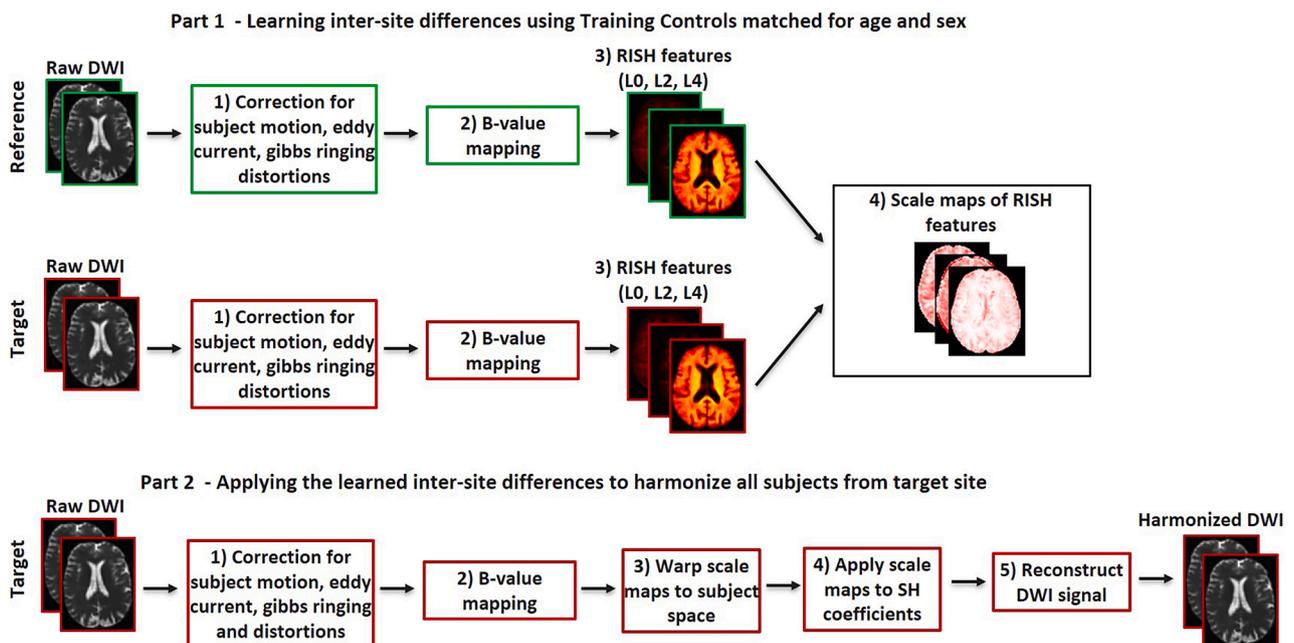


Fig. 1. Harmonization steps using RISH features. Part 1) All scans are pre-processed to correct for artefacts, followed by b-value mapping to a common b-value of 1000 s/mm^2 . Voxel-wise scale maps are computed using a set of Training Controls matched for age and sex between the reference and the target site. Part 2) The scale maps are then applied to harmonize the remaining scans of the target site.

(demographics in Table 1). We made a selection of controls that included both training and validation controls, while ensuring that the average age was similar across sites (demographics in supplementary Table S3). We matched controls for age across sites in order to have a common reference to calculate effect sizes against patients. Thus, after harmonization controls are expected to have similar diffusion measures while the contrast with their respective patient groups should not be affected. To assess sensitivity to disease effects, general linear model (GLM) adjusted for age and sex was performed to examine the contrast (effect sizes of FA, MD and PSMD from the TBSS skeleton) between patients and controls within each site, before and after harmonization.

In patients, we also explored the sensitivity to disease effects by relating WMH volume with FA, MD and PSMD, before and after harmonization. Linear regression adjusted for age and sex was performed. Since the WMH volumes was non-normally distributed, a Box-Cox transformation was applied (Box & Cox, 1964). We determined the R^2 and standardized regression coefficients (β), before and after harmonization.

2.4.3. Proof of concept of data pooling

We evaluated if disease effects were similar if patients were compared to Internal Controls or to a pooled set of matched controls derived from external centres only. GLM adjusted for age and sex was performed to compare patients from Utrecht1 and from Munich versus External Controls pooled from other sites. We compared effect sizes of FA, MD, PSMD obtained with External Controls to the original effect sizes with Internal Controls.

Finally, we demonstrated proof of concept of data pooling by relating WMH volume with FA, MD and PSMD on pooled data of patients from multiple sites, and compared the fit of the curve before and after harmonization. Similar to the analysis within sites, linear regression was performed.

3. Results

3.1. Effectiveness of RISH harmonization

Fig. 2 shows RISH features of order 0, 2 and 4 obtained from the Training Controls and corresponding scale maps, before harmonization.

We observed widespread differences in RISH features between the reference and target sites that were dependent on the tissue type. Small differences were observed in regions with prevalently single fiber populations (e.g., corpus callosum) for all orders of RISH features, whereas more peripheral white matter and grey matter regions showed bigger differences across sites (see scale maps). Furthermore, data from Hong Kong, Munich and Singapore sites showed larger differences from the reference (Utrecht1) than Utrecht2.

Before harmonization, significant differences in FA and MD were found between the reference and each target site across the entire brain (Fig. 3), especially for Hong Kong Munich and Singapore ($p < 0.05$). After harmonization, all significant differences in the white matter skeleton between the target sites and the reference were removed. When analysing the whole brain, FA differences were still seen for the Hong Kong site ($p < 0.05$), mainly in subcortical grey matter and near tissue interfaces with cerebrospinal fluid, probably due to misregistration (Fig. 3, top right panel). A map of effect sizes further clarifies that differences (positive or negative Cohen's d) are removed after harmonization (i.e., effect sizes become closer to zero, Supplementary Figure S1).

Regarding Validation Controls (Fig. 4), voxel-wise differences in FA and MD in the white matter skeleton and across the whole brain were removed after harmonization, with exception of minor differences at tissue interfaces, probably due to misregistration. When comparing the average FA and MD of the skeleton, significant differences in FA were found across sites before harmonization ($F(1,43) = 18.2$, $p < 0.001$, Fig. 5A). All differences in FA were removed after harmonization. The average MD of the skeleton did not differ across sites before or after harmonization (Fig. 5B).

3.2. Sensitivity to disease effects in SVD

Fig. 6 depicts differences in dMRI metrics between patients and controls within each site, with quantitative values shown in Table 2. Before harmonization, patients had a significantly lower FA (Fig. 6A), higher MD (Fig. 6B) and higher PSMD (Fig. 6C) than controls in all sites except Hong Kong: FA ($d = -0.96$ to -2.07 , $p < 0.001$); MD ($d = 1.02$ to 1.99 , $p < 0.001$); PSMD ($d = 0.93$ to 1.71 , $p < 0.001$). After harmonization, all effect sizes were preserved, regardless of if they were small (0.2), medium (0.5) or large (0.8). On average, the relative change in

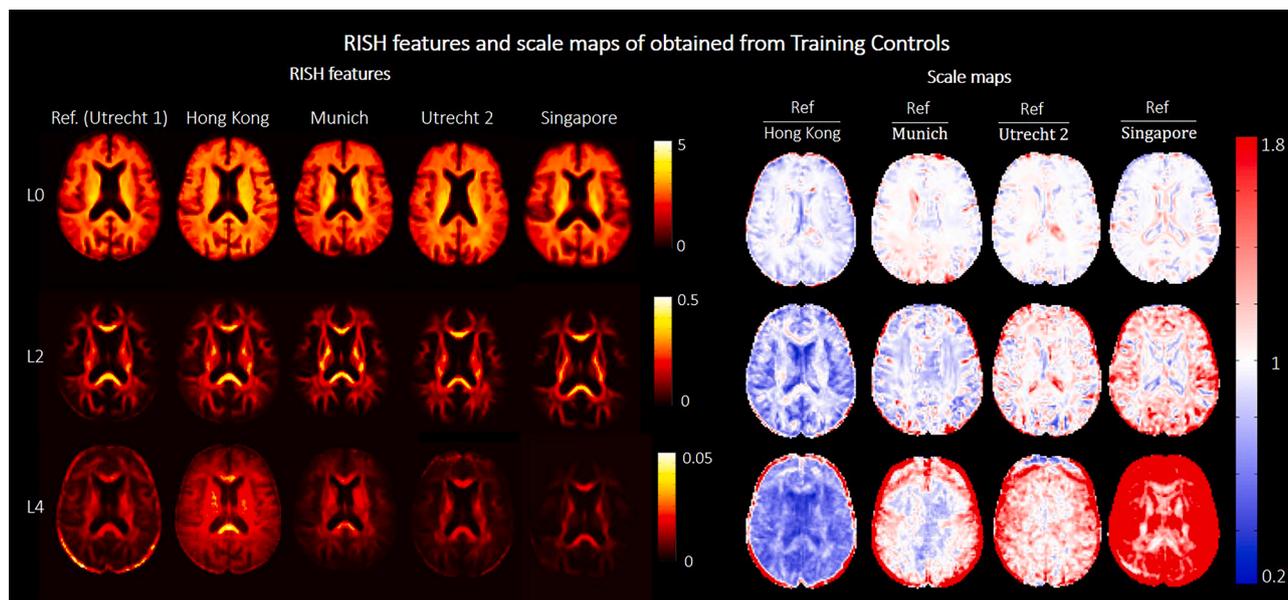


Fig. 2. Left (top to bottom): RISH features of order 0, 2 and 4 calculated with Training Controls of each site. Different columns correspond to the different sites. Right (top to bottom): Scale maps for each RISH feature obtained by scaling each target site to the reference. Different columns correspond to different target sites. The red-blue colormap indicates the scaling factor between the two sites. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

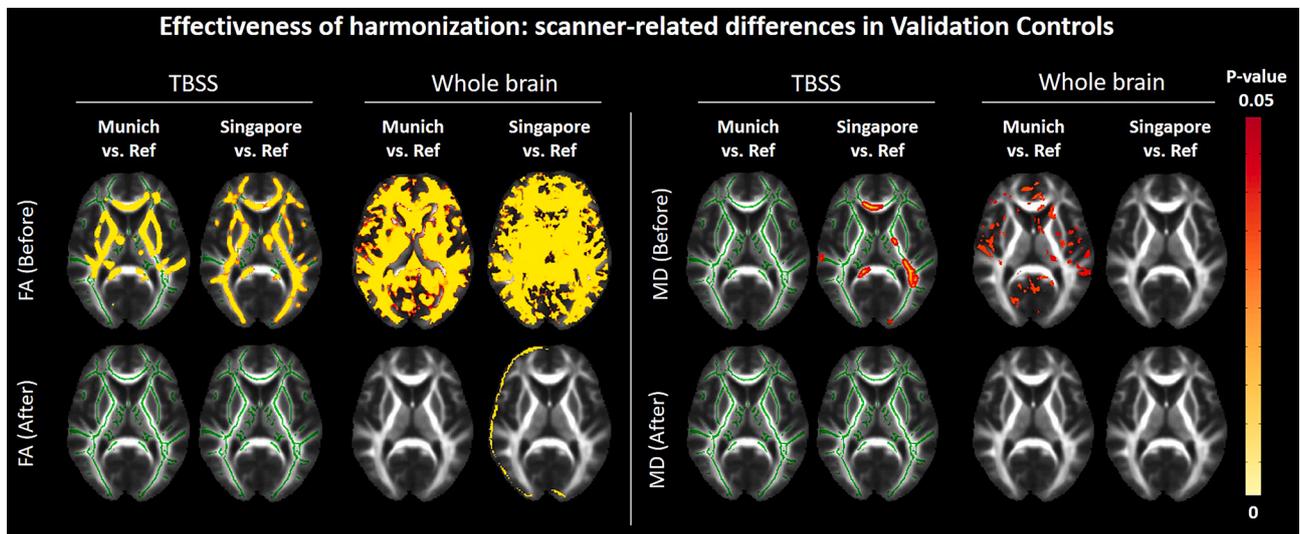


Fig. 4. Results of the TBSS and whole brain voxel-wise analysis comparing FA (left) and MD (right) between Validation Controls of each target site and the reference, before and after harmonization (demographics in supplementary Table S2). The yellow–red colormap shows voxels where statistical differences were observed after multiple comparison corrections (p -value < 0.05). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

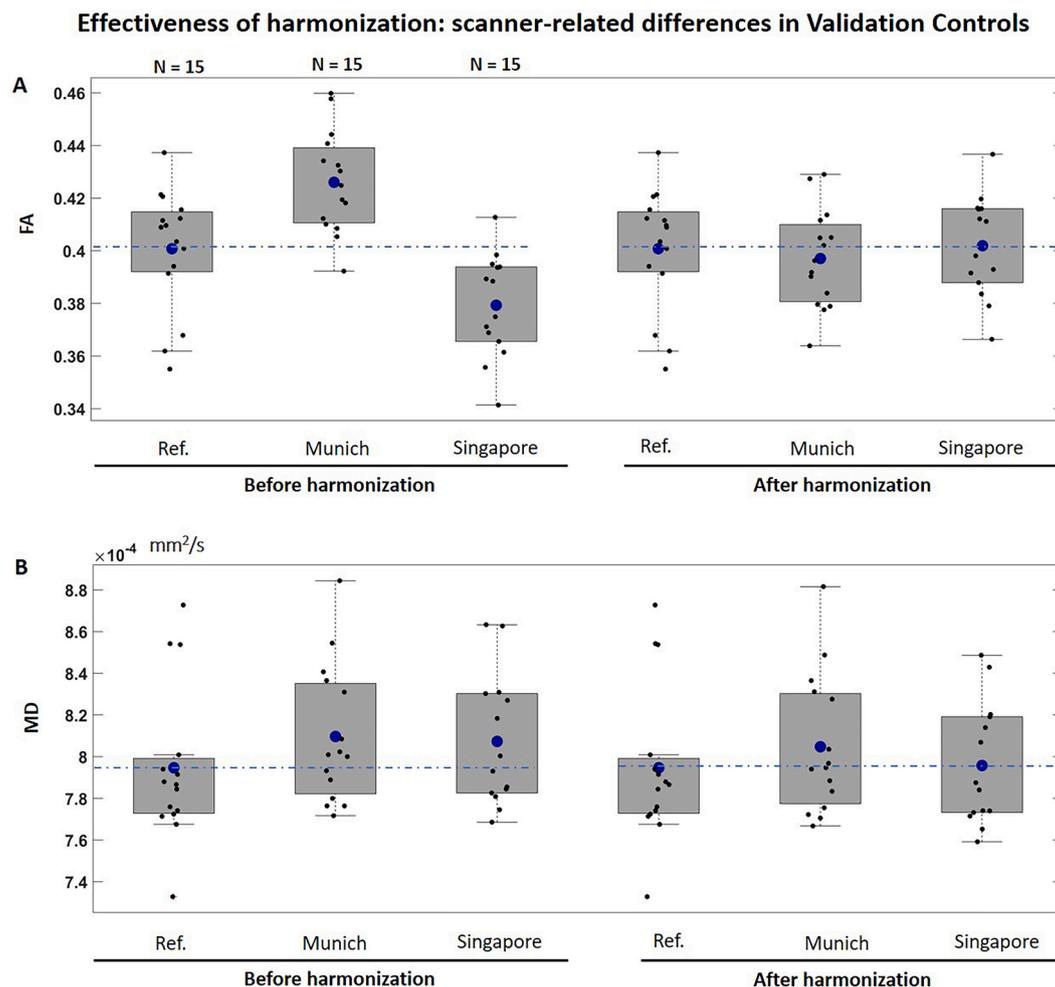


Fig. 5. Boxplots of average FA (panel A) and MD (panel B) of the white matter skeleton compared between Validation Controls, before and after harmonization (demographics in supplementary Table S2). The dashed blue line represents the mean of the reference site. The blue marker within each boxplot indicates the mean of the corresponding group, which were compared between sites using a one-way ANOVA. Top: FA before harmonization, $F(1,43) = 18.2$, $p < 0.001$; FA after harmonization, $F(3,43) = 0.2$, $p = 0.8$. Bottom: MD before harmonization: $F(3,43) = 0.83$, $p = 0.06$; MD after harmonization: $F(3,43) = 2.6$, $p = 0.4$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

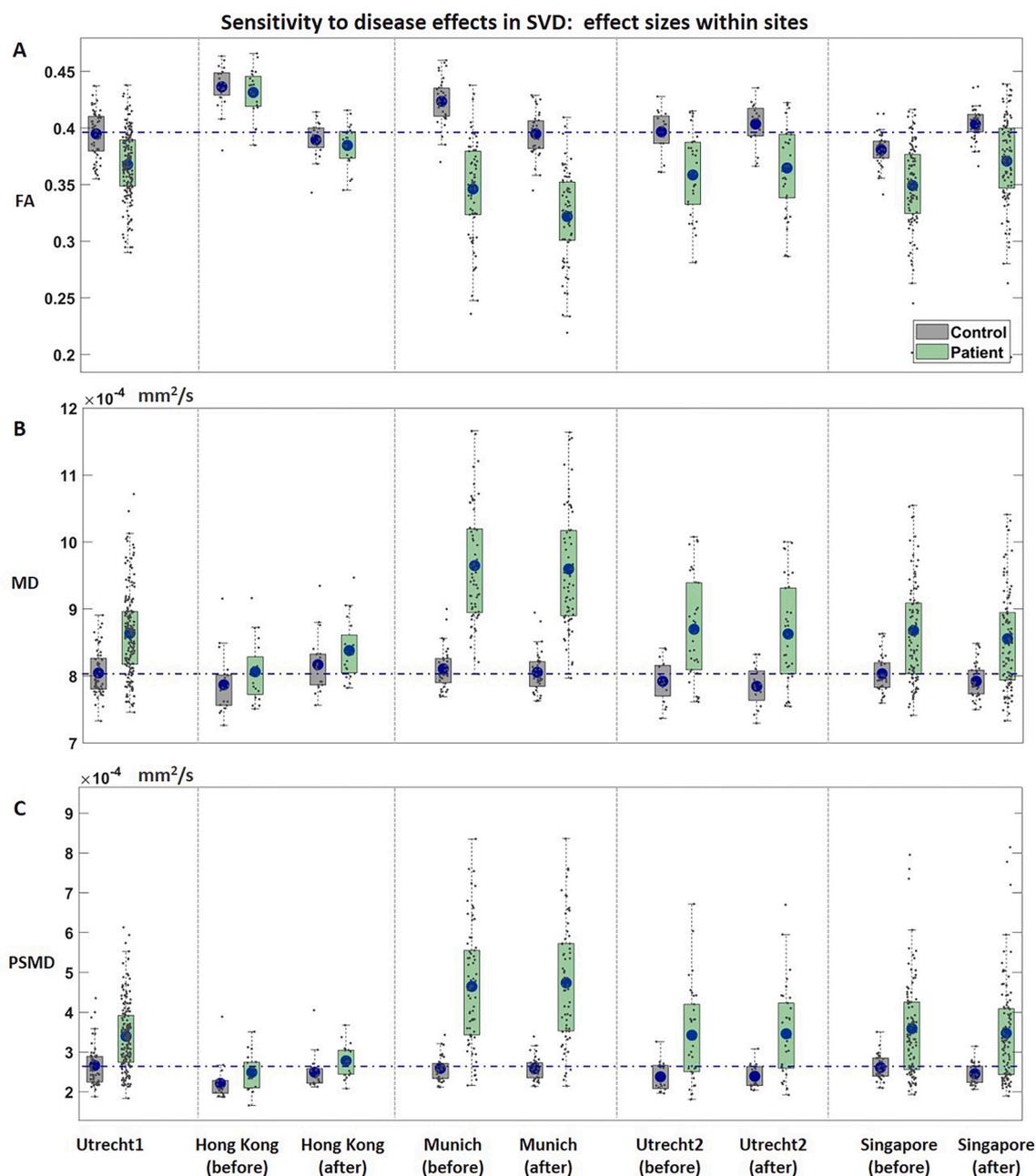


Fig. 6. Average FA (panel A), MD (panel B) and PSMD (panel C) of the white matter skeleton compared between patients (green) and controls (gray) within each site (demographics for selected controls in supplementary Table S3). Results are displayed for the reference site and for each target site before and after harmonization. Corresponding p-values and effect sizes are displayed in Table 2. The dashed line indicates the mean value of controls of the reference site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

We investigated the applicability of RISH harmonization to remove acquisition-related differences in multicentre dMRI of elderly subjects with SVD while preserving disease-related effects. Before harmonization, we observed significant differences in FA and MD across sites, which were removed after harmonization, both in the Training Controls and in Validation Controls not involved in the training step. Importantly, effect sizes of FA, MD and PSMD for group differences between patients and controls as well as for associations with WMH volume within each site were preserved after harmonization. The harmonized controls could be effectively considered as a single-site dataset. The pooled data of patients covered a wide range of WMH burden, allowing to demonstrate a strong relation between WMH volume and dMRI metrics.

The RISH method has been previously implemented using scans of healthy young subjects without apparent brain lesions for the training step (Cetin-Karayumak et al., 2020; Mirzaalian et al., 2015). Here, we have evaluated whether peculiar characteristics of the elderly brain, such as presence of lesions as WMH brain atrophy and larger ventricles, which are present to some extent even in control subjects, could affect the computation of the scale maps of RISH features. In our study, the selection criteria for controls was based on low burden of SVD. Thus, the training controls were minimally affected by WMH and had relatively similar brain volumes. The largest differences in RISH features were observed in grey matter and peripheral white matter areas, where partial volume effects might play a role on the diffusion profile (Vos et al., 2011, 2012). After harmonization differences in FA and MD between the target sites and the reference were removed across the brain, except for

Table 2
Effect sizes of FA, MD and PSMD between patients and controls within sites, before and after harmonization.

	Utrecht1	Hong Kong before	Hong Kong after	% Change	Munich before	Munich after	% Change	Utrecht 2 before	Utrecht 2 after	% Change	Singapore before	Singapore after	% Change
FA	1.40×10^{-11}	0.43	0.39	8	3.17×10^{-21}	2.30×10^{-21}	1	2.20×10^{-5}	2.20×10^{-5}	0	3.30×10^{-10}	9.00×10^{-15}	2
P-value	0.96	0.25	0.27		-2.07	-2.09		-1.16	-1.16		-1.19	-1.17	
Effect size (Cohen's d)													
CI	(-1.29, -0.64)	(-0.87, 0.37)	(-0.88, 0.35)		(-2.57, -1.57)	(-2.59, -1.59)		(-1.79, -0.51)	(-1.79, -0.52)		(-1.54, -0.83)	(-1.53, -0.81)	
MD	6.29×10^{-10}	0.17	0.13	9	9.57×10^{-21}	6.29×10^{-21}	2	2.84×10^{-4}	2.62×10^{-4}	1	2.40×10^{-14}	5.80×10^{-14}	3
P-value	1.02	0.44	0.48		1.99	2		1.18	1.19		1.14	1.11	
Effect size (Cohen's d)													
CI	(0.63, 1.34)	(-0.19, 1.06)	(-0.15, 1.11)		(1.49, 2.49)	(1.49, 2.48)		(0.54, 1.82)	(0.55, 1.83)		(0.80, 1.50)	(0.76, 1.50)	
PSMD	3.10×10^{-12}	0.06	0.04	11	3.6×10^{-17}	5.97×10^{-18}	4	3.2×10^{-6}	9×10^{-6}	7	4.10×10^{-12}	2.40×10^{-12}	0
P-value	0.93	0.62	0.69		1.71	1.78		1.03	1.1		-0.99	-0.99	
Effect size (Cohen's d)													
CI	(0.62, 1.25)	(-0.02, 1.25)	(0.05, 1.32)		(1.23, 2.18)	(1.29, 2.25)		(0.39, 1.65)	(0.46, 1.73)		(-1.34, -0.64)	(-1.35, -0.65)	

Demographics for selected controls in Supplementary Table S3.

Hong Kong where minor differences in MD still persisted in deep grey matter structures and at tissue interfaces with cerebrospinal fluid. This is likely due to residual inaccuracies in image registration or differences in WMH burden and brain volumes in Training controls, given these were not explicitly matched for these markers (Supplementary Figure S7). Accordingly, we suggest that when dealing with data of elderly subjects it might be beneficial to match Training Controls not only for age and sex (Hsu et al., 2008), but also in terms of WMH lesion distribution, brain volumes or other demographics that contribute for variation in diffusion (e.g., handedness, race, etc., Büchel et al., 2004), although this might be challenging to achieve in practice in most studies. This is particularly important when the inclusion criteria for controls are not based on MRI markers of SVD but rather on variables such as being cognitively healthy. Another important consideration for studies implementing the RISH method is that imaging artefacts specific of one site or few subjects (e.g., ghosting, incomplete fat saturation) might be learned as part of the harmonization features, and propagate into the harmonized dataset. In our study, an example of these artefacts can be seen as rings due to incomplete fat suppression on the L4 scale maps shown in Fig. 2. Nevertheless, their impact on the harmonized data was deemed minimal and it did not significantly affect any subsequent result. The RISH method does not assume that two groups of healthy older subjects are completely identical, but as shown by our work, if groups are matched for major factors, differences in RISH features can be learned on a group-level, without major influence of individual properties. This is further supported by our results on the generalizability of RISH harmonization with Validation Controls. We demonstrated not only that dMRI metrics of subjects not involved in the training step are harmonized, but also the transitivity of harmonization, e.g., that the independent harmonization of two target sites to the reference also implies harmonization between target sites.

Next, we demonstrated that RISH harmonization does not affect the sensitivity of dMRI to effects of SVD. Well-known differences in FA, MD and PSDM between patients and controls observed within each site before harmonization were preserved (Baykara et al., 2016; Wiegertjes et al., 2019). Effect sizes were unaffected after harmonization (relative change = 3.9 %), regardless of whether the magnitude was small, medium or large. We believe such small change in effect size can be deemed negligible, and is likely caused by registration and interpolation inaccuracies (Karayumak et al., 2019). We also repeated the same analysis with the statistical harmonization method ComBat (Fortin et al., 2017) for comparison. RISH harmonization outperformed ComBat in this dataset, which was not able to preserve effect sizes within all sites (see supplementary Information, Part 6). Recent work has shown that application of the RISH method does not alter the relation between dMRI metrics and biological effects such as age-related changes (Karayumak et al., 2019; 2020). Here, we extended such finding by showing that RISH harmonization also preserved the relation between WMH volume and dMRI metrics, a well-established relation in this kind of patients (Van Leijssen et al., 2017), regardless of the strength or the sample size available within each site to test such associations.

To date, most studies of SVD with dMRI have been single-site based, and the inclusion of cohorts from other sites, which can differ substantially in terms of acquisition, has been limited to external validation only. Our results with External Controls indeed show that multicentre data without standardized acquisition across centres cannot be simultaneously analysed, as their integration before harmonization would result in biased effect sizes when comparing patients and controls. Effect sizes obtained with External Controls before harmonization were biased up to 1 standard deviation, which is on the same order of magnitude as typical differences between patients and controls. Conversely, after harmonization External Controls behaved as single-site dataset that could be used as reference for patients from all sites with minimal bias in effect sizes. An important implication of this result is that harmonization can potentially address data obsolescence. Diffusion scans are routinely acquired at single-site level for the purpose of testing specific

Sensitivity to disease effects in SVD: Associations with WMH volume within sites

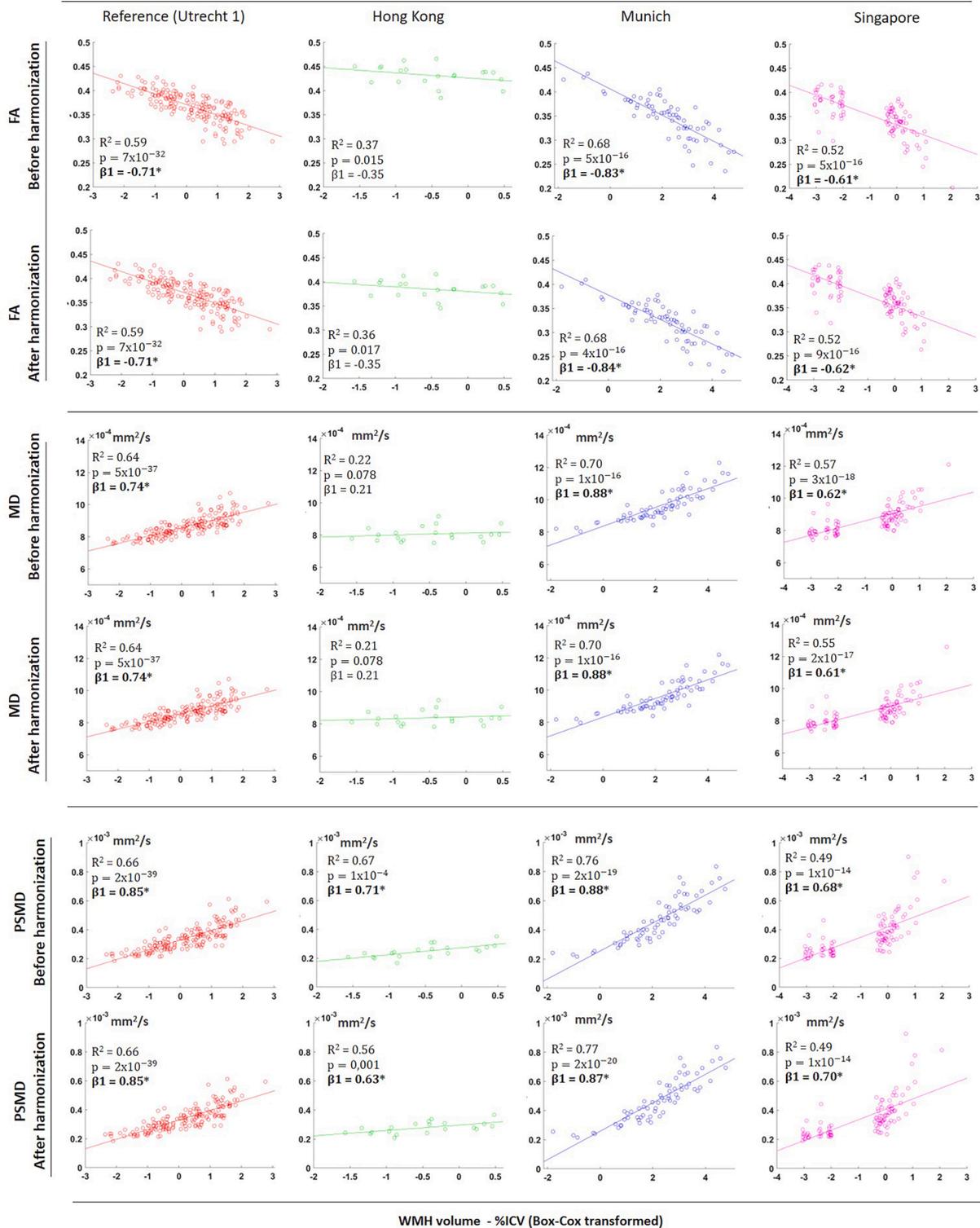


Fig. 7. Scatter plots of associations between WMH volume, FA (top), MD (middle) and PSMD (bottom) within sites, before and after harmonization. MD and PSMD values are given in mm^2/s . * $p < 0.05$ for β coefficients.

hypotheses, and discarded afterwards as hardware updates are implemented or the acquisition protocols are adjusted. Being able to account for such differences might allow to include previously acquired data across multiple studies, thus valorising previous investments and reducing the burden of scanning new controls in prospective studies. Another potential benefit of harmonization is in longitudinal studies

where upgrades in scanner systems complicate the comparison of data at different time points (Takao et al., 2012). Still, for prospective multi-center studies and especially clinical trials, which typically lack a healthy control group needed for post-hoc data harmonization, standardization of the acquisition should still have a high priority.

After establishing that RISH harmonization allows to integrate data

Proof of concept of data pooling: effect sizes using External Controls

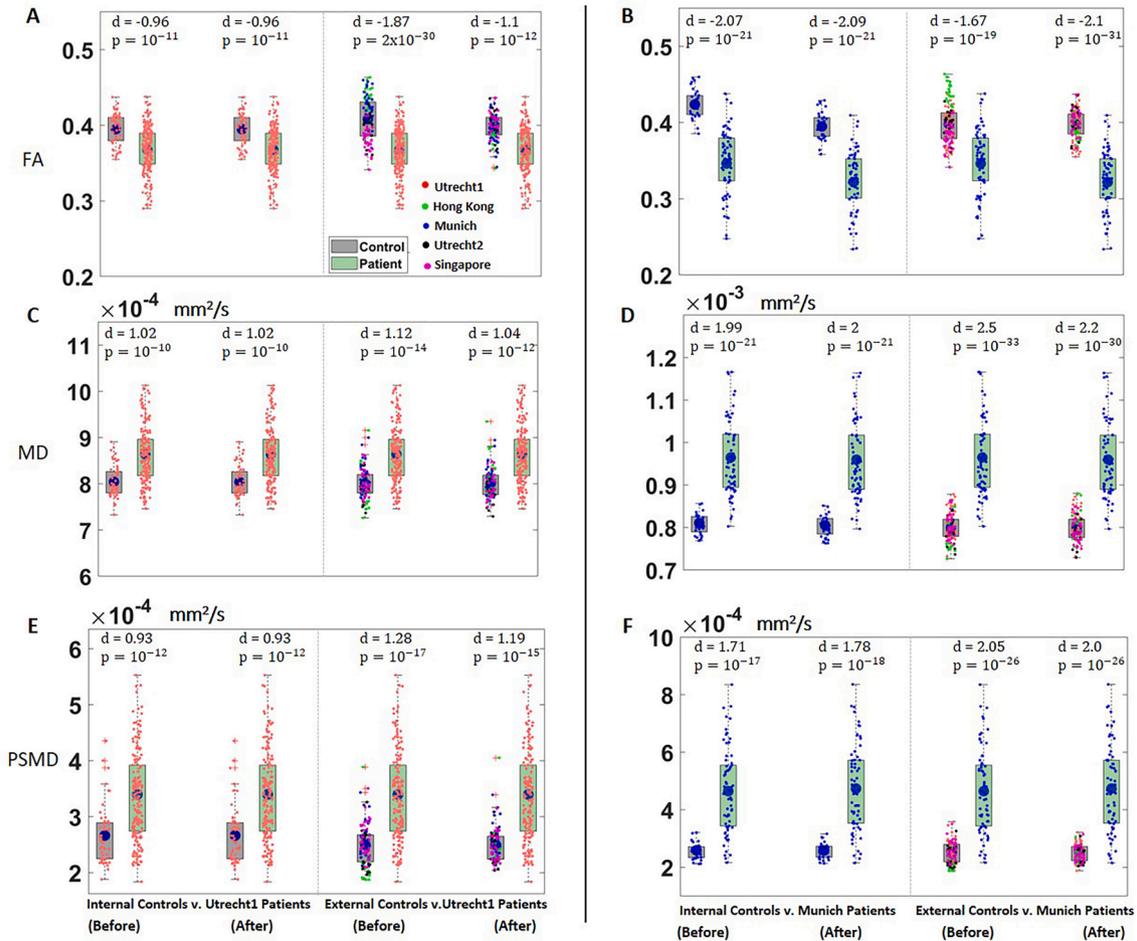


Fig. 8. Boxplots of FA (A-B), MD (C-D) and PSMD (E-F) of the white matter skeleton compared between patients and controls. (Left): Comparisons between Utrecht1 patients and Internal Controls and between Utrecht1 patients and External Controls, before and after harmonization. (Right): Comparisons between Munich patients and Internal Controls, and between Munich patients and External Controls, before and after harmonization (demographics for selected controls in supplementary Table S3). Data from each site is color coded as follows: red = Utrecht1; green = Hong Kong; blue = Munich; black = Utrecht2; magenta = Singapore. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from different cohorts, we demonstrated associations between WMH volume and dMRI metrics on the pooled dataset of patients, which resulted into improved statistical power due to the larger sample size. Since the cohorts included in this study had different disease burdens, the pooled data covered a larger spectrum of WMH volumes, allowing to test associations with more confidence than what the individual sites would allow. Since WMH volume already has strong correlations with dMRI metrics (Bendlin et al., 2010), even the non-harmonized pooled data resulted in significant associations. However, R^2 were lower than some individual sites, showcasing again the risk of performing inferences using non-harmonized data. After harmonization, data points of different sites behaved as a single-center data and more aligned with the fitted curve, resulting in an increase of R^2 . The impact of harmonization is likely to be even more important in the study of other correlates of SVD with more subtle effect sizes than WMH (e.g., relation between dMRI metrics and cognition, Du et al., 2020) or when assessing disease progression over time (van Leijssen et al., 2019).

5. Limitations and future directions

Despite its advantages, RISH harmonization does not come without limitations. To minimize differences in diffusion weighting across sites, we used a linear scaling to map the diffusion signal to a b-value of 1000

s^2/mm , which is only applicable to a limited range of b-values. For prospective multicentre studies, it is crucial to minimize differences in acquisition parameters such as b-values, but for retrospective studies with already acquired data, harmonization of scans with largely different b-values needs further investigation. As clinical protocols continuously improve and more complex sequences are implemented (e.g., multi-shell data), future work should also investigate whether RISH harmonization is suitable for such advanced dMRI applications in SVD (De Luca et al., 2018; Konieczny et al., 2020; Rydhög et al., 2017). Moreover, since we focused our analyses on the dMRI metrics most commonly associated with SVD (FA, MD, PSMD), further analyses are required in order to generalize our conclusions to other metrics obtained from higher level analysis such as fiber tractography (De Luca et al., 2020) and network theory in SVD (Reijmer et al., 2015).

6. Conclusions

Despite the limitations, our study is the first to prove the feasibility of RISH harmonization of multicentre dMRI scans in the context of SVD. We showed that harmonizing the raw dMRI signal is effective in removing acquisition-related differences, while preserving the sensitivity to disease effects. This ultimately allowed us to directly pool scans acquired at different sites into a single analysis and increase the power of

Proof of concept of data pooling: Associations between WMH volume and dMRI metrics

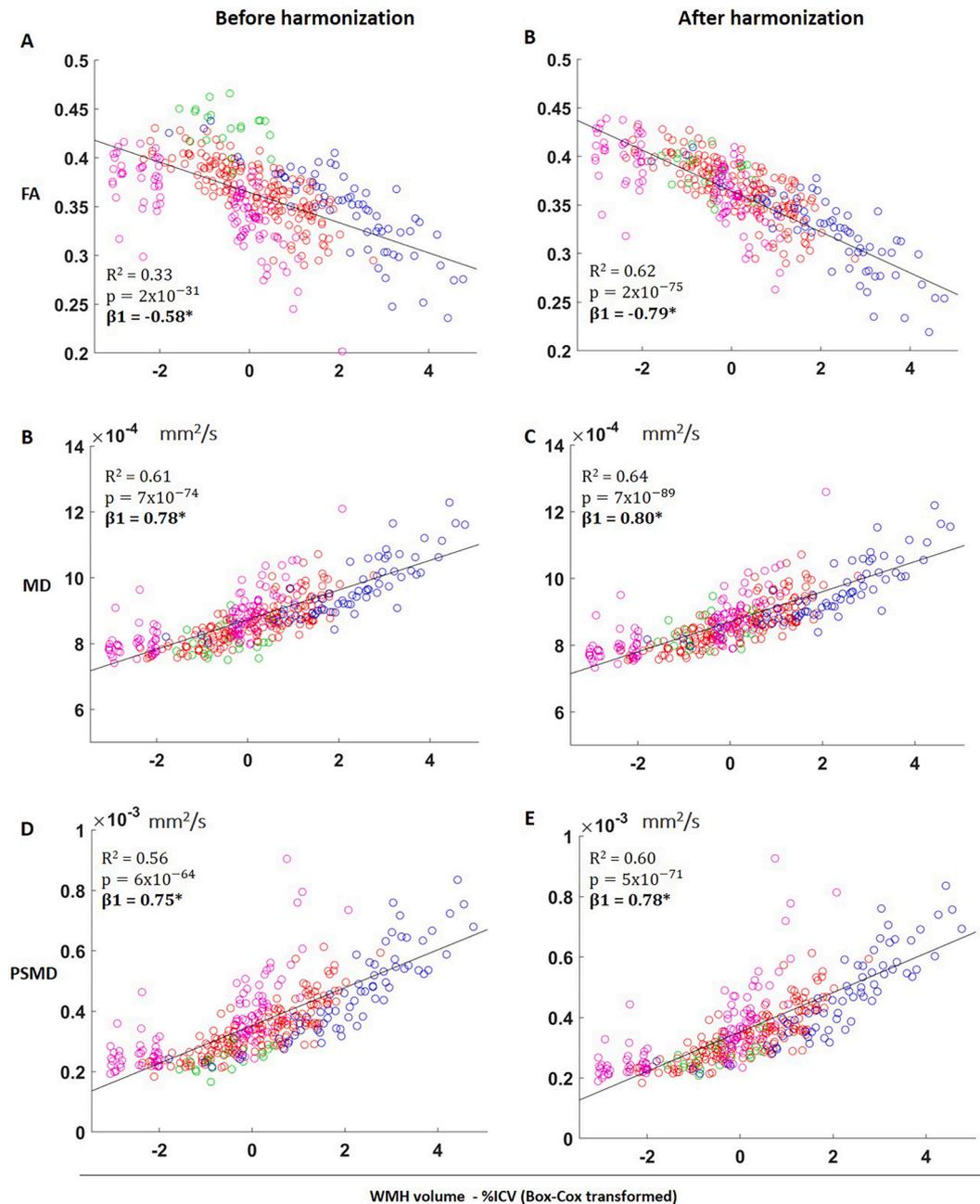


Fig. 9. Scatter plots of associations between WMH volume, FA (top), MD (middle) and PSMD (bottom) on the pooled data, before and after harmonization. MD and PSMD values are given in mm²/s. * $p < 0.05$ for β coefficients. Data from each site is color coded as follows: red = Utrecht1; green = Hong Kong; blue = Munich; magenta: Singapore. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dMRI inferences. Our work paves the way not only for the validation of dMRI markers of SVD in large scale multicentre studies, but also for studies that aim to answer new research questions where statistical power is critical, such as untangling underlying aetiologies in SVD populations (e.g., free-water imaging, Finsterwalder et al., 2020). When translated to even larger scales, harmonization could significantly improve the sensitivity and specificity for studies attempting to identify tract specific damage and network connections related to cognitive dysfunction (Konieczny et al., 2020).

7. Data availability statement

The data used in this study is available upon contact and agreement with the respective investigators of each cohort. The harmonization method used in this work was originally developed by Mirzaalian et al., 2015 and the code is publicly available at <https://github.com/pnlbwh/dMRIharmonization>.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Study funding

This work was supported by ZonMw, The Netherlands Organisation for Health Research and Development (VICI grant 91816616 to G.J. Biessels). The research of A. Leemans is supported by VIDI Grant 639.072.411 from the Netherlands Organization for Scientific Research (NWO). Zoom@SVDs is part of the SVDs@target project. SVDs@target has received funding from the European Union's Horizon2020 research and innovation program under grant agreement No 666881. The CUHK-RI is supported by General Research Fund (grant number GRF CUHK 471911), the Lui CheWoo Institute of Innovative Medicine, and Therese Pei Fong Chow Research Centre for Prevention of Dementia (in memory of Donald H. K. Chow). Funding for the EDIS study was provided by the National Medical Research Council of Singapore.

Acknowledgements

We acknowledge and thank all patients and controls for study participation. We thank members of the Utrecht Vascular Cognitive Impairment (VCI) Study group. We also thank Mathias Hübner for technical assistance and Angelika Doerr as study nurse for her help in the VASCAMY study in Munich. We acknowledge and thank all patients and controls for study participation. We thank members of the Utrecht Vascular Cognitive Impairment (VCI) Study group. We also thank Mathias Hübner for technical assistance and Angelika Doerr as study nurse for her help in the VASCAMY study in Munich.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2021.102886>.

References

- Aalten, P., Ramakers, I.H.G.B., Biessels, G.J., de Deyn, P.P., Koek, H.L., OldeRikkert, M. G.M., Oleksik, A.M., Richard, E., Smits, L.L., van Swieten, J.C., Teune, L.K., van der Lugt, A., Barkhof, F., Teunissen, C.E., Rozendaal, N., Verhey, F.R.J., van der Flier, W. M., 2014. The Dutch Parelnoer Institute - Neurodegenerative diseases; methods, design and baseline results. *BMC Neurol.* 14 (1), 1–8. <https://doi.org/10.1186/s12883-014-0254-4>.
- Avants, Brian B., Tustison, Nicholas J., Song, Gang, Cook, Philip A., Klein, Arno, Gee, James C., et al., 2010. A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration. *Neuroimage* 54, 2033–2044. <https://doi.org/10.1016/j.neuroimage.2010.09.025>.
- Baykara, E., Gesierich, B., Adam, R., Tuladhar, A.M., Biesbroek, J.M., Koek, H.L., Ropele, S., Jouvent, E., Chabriat, H., Ertl-Wagner, B., Ewers, M., Schmidt, R., de Leeuw, F.-E., Biessels, G.J., Dichgans, M., Duering, M., 2016. A novel imaging marker for small vessel disease based on skeletonization of white matter tracts and diffusion histograms. *Ann. Neurol.* 80 (4), 581–592. <https://doi.org/10.1002/ana.24758>.
- Bendlin, B.B., Fitzgerald, M.E., Ries, M.L., Xu, G., Kastman, E.K., Thiel, B.W., Rowley, H. A., Lazar, M., Alexander, A.L., Johnson, S.C., 2010. White matter in aging and cognition: A cross-sectional study of microstructure in adults aged eighteen to eighty-three. *Dev. Neuropsychol.* 35 (3), 257–277. <https://doi.org/10.1080/87565641003696775>.
- Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 26 (2), 211–243. <https://doi.org/10.1111/rssb.1964.26.issue-210.1111/j.2517-6161.1964.tb00553.x>.
- Brink, H. Van Den, Kopczak, A., Arts, T., Onkenhout, L., Siero, J. C. W., Zwanenburg, J. J. M., Duering, M., Blair, G. W., Doubal, F. N., Stringer, M. S., Thrippleton, M. J., Kuijff, H. J., Luca, A. De, Hendrikse, J., Wardlaw, J. M., Dichgans, M., & Jan, G. (2021). Cerebral Circulation - Cognition and Behavior Zooming in on cerebral small vessel function in small vessel diseases with 7T MRI : Rationale and design of the “ ZOOM @ SVDs ” study. *Cerebral Circulation - Cognition and Behavior*, 2(April), 100013. <https://doi.org/10.1016/j.cccb.2021.100013>.
- Büchel, C., Raedler, T., Sommer, M., Sach, M., Weiller, C., Koch, M.A., 2004. White matter asymmetry in the human brain: A diffusion tensor MRI study. *Cereb. Cortex* 14 (9), 945–951. <https://doi.org/10.1093/cercor/bhh055>.
- Cetin-Karayumak, S., Di Biase, M.A., Chunga, N., Reid, B., Somes, N., Lyall, A.E., Kelly, S., Solgun, B., Pasternak, O., Vangel, M., Pearson, G., Tamminga, C., Sweeney, J.A., Clementz, B., Schretlen, D., Viher, P.V., Stegmayer, K., Walther, S., Lee, J., Crow, T., James, A., Voinikos, A., Buchanan, R.W., Szeszko, P.R., Malhotra, A.K., Hegde, R., McCarley, R., Keshavan, M., Shenton, M., Rath, Y., Kubicki, M., 2020. White matter abnormalities across the lifespan of schizophrenia: a harmonized multi-site diffusion MRI study. *Mol. Psychiatry* 25 (12), 3208–3219. <https://doi.org/10.1038/s41380-019-0509-y>.
- 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 (January), 117206. <https://doi.org/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.* 31 (11), 1–17. <https://doi.org/10.1002/nbm.3965>.
- Du, J., Zhu, H., Zhou, J., Lu, P., Qiu, Y., Yu, L., Cao, W., Zhi, N., Yang, J., Xu, Q., Sun, J., Zhou, Y., 2020. Structural Brain Network Disruption at Preclinical Stage of Cognitive Impairment Due to Cerebral Small Vessel Disease. *Neuroscience* 449, 99–115. <https://doi.org/10.1016/j.neuroscience.2020.08.037>.
- Fazekas, F., Chawluk, J.B., Alavi, A., 1987. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *Am. J. Neuroradiol.* 8 (3), 421–426.
- Finstervwalder, S., Vlegels, N., Gesierich, B., Araque Caballero, M., Weaver, N. A., Franzmeier, N., Georgakis, M. K., Konieczny, M. J., Koek, H. L., Karch, C. M., Graff-Radford, N. R., Salloway, S., Oh, H., Allegri, R. F., Chhatwal, J. P., Jessen, F., Düzel, E., Dobisch, L., Metzger, C., ... Duering, M. (2020). Small vessel disease more than Alzheimer's disease determines diffusion MRI alterations in memory clinic patients. *Alzheimer's and Dementia, March*, 1504–1514. <https://doi.org/10.1002/alz.12150>.
- Fortin, J.P., Parker, D., Tunc, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T., 2017. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161 (July), 149–170. <https://doi.org/10.1016/j.neuroimage.2017.08.047>.
- Helmer, K.G., Chou, M.-C., Preciado, R.I., Gimi, B., Rollins, N.K., Song, A., Turner, J., Mori, S., 2016. Multi-site study of diffusion metric variability: effects of site, vendor, field strength, and echo time on regions-of-interest and histogram-bin analyses. *Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging* 9788, 97882U. <https://doi.org/10.1117/12.2217445>.
- Hilal, S., Ikram, M.K., Saini, M., Tan, C.S., Catindig, J.A., Dong, Y.H., Lim, L.B.S., Ting, E. Y.S., Koo, E.H., Cheung, C.Y.L., Qiu, A., Wong, T.Y., Chen, C.L.H., Venketasubramanian, N., 2013. Prevalence of cognitive impairment in Chinese: Epidemiology of Dementia in Singapore study. *J. Neurol. Neurosurg. Psychiatry* 84 (6), 686–692. <https://doi.org/10.1136/jnnp-2012-304080>.
- Hsu, J.L., Leemans, A., Bai, C.H., Lee, C.H., Tsai, Y.F., Chiu, H.C., Chen, W.H., 2008. Gender differences and age-related white matter changes of the human brain: A diffusion tensor imaging study. *Neuroimage* 39, 566–577. <https://doi.org/10.1016/j.neuroimage.2007.09.017>.
- Jensen, Jens H., Helpert, Joseph A., Ramani, Anita, Lu, Hanzhang, Kaczynski, Kyle, et al., 2005. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine* 53 (6), 1432–1440. <https://doi.org/10.1002/mrm.20508>.
- Karayumak, S., Bouix, S., Ning, L., James, A., Crow, T., Shenton, M., Kubicki, M., Rath, Y., 2019. Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. *NeuroImage* 184, 180–200. <https://doi.org/10.1016/j.neuroimage.2018.08.073>.
- 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., 2020. Multi-shell diffusion MRI models for white matter characterization in cerebral small vessel disease. In *Neurology*. <https://doi.org/10.1212/wnl.0000000000011213>.
- Koppers, S., Bloy, L., Berman, J.I., Tax, C.M.W., Edgar, J.C., Merhof, D., 2019. Spherical Harmonic Residual Network for Diffusion Signal Harmonization. *Mathematics and Visualization* 226249, 173–182. https://doi.org/10.1007/978-3-030-05831-9_14.
- Kuijff, H.J., et al., 2019. Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities and Results of the WMH Segmentation Challenge. *IEEE Transactions on Medical Imaging* 38 (11), 2556–2568. <https://doi.org/10.1109/TMI.2019.2905770>.
- Kumar, R., Husain, M., Gupta, R.K., Hasan, K.M., Harris, M., Agarwal, A.K., Pandey, C.M., Narayana, P.A., 2009. Serial changes in the white matter diffusion tensor imaging metrics in moderate traumatic brain injury and correlation with neuro-cognitive function. *J. Neurotrauma* 26 (4), 481–495. <https://doi.org/10.1089/neu.2008.0461>.
- Lam, B.Y.K., Leung, K.T., Yiu, B., Zhao, L., Biesbroek, J.M., Au, L., Tang, Y., Wang, K., Fan, Y., Fu, J., Xu, Q., Song, H., Tian, X., Chu, W.C.W., Abrigo, J., Shi, L., Ko, H., Lau, A., Duering, M., Wong, A., Mok, V.C.T., 2019. Peak width of skeletonized mean diffusivity and its association with age-related cognitive alterations and vascular risk factors. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring* 11 (1), 721–729. <https://doi.org/10.1016/j.dadm.2019.09.003>.
- van Leijssen, E.M.C., van Uden, I.W.M., Bergkamp, M.I., van der Holst, H.M., Norris, D.G., Claassen, J.A.H.R., Kessels, R.P.C., de Leeuw, F.-E., Tuladhar, A.M., 2019. NeuroImage : Clinical Longitudinal changes in rich club organization and cognition in cerebral small vessel disease. *NeuroImage: Clinical* 24, 102048. <https://doi.org/10.1016/j.nicl.2019.102048>.
- Lorca-Puls, D. L., Gajardo-Vidal, A., White, J., Seghier, M. L., Leff, A. P., Green, D. W., Crinion, J. T., Ludersdorfer, P., Hope, T. M. H., Bowman, H., & Price, C. J. (2018). The impact of sample size on the reproducibility of voxel-based lesion-deficit mappings. *Neuropsychologia*, 115(June 2017), 101–111. <https://doi.org/10.1016/j.neuropsychologia.2018.03.014>.
- Leemans, Alexander, Jeurissen, Ben, Sijbers, Jan, et al., 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. *Proceedings of the International Society for Magnetic Resonance in Medicine*. <https://doi.org/10.1093/ocmed/kqr069>.

- Leemans, Alexander, Jones, Derek K., 2009. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magnetic Resonance in Medicine* 61 (6), 1336–1349. <https://doi.org/10.1002/mrm.21890>.
- Lyoubi-Idrissi, A.L., Jouvent, E., Poupon, C., Chabriat, H., 2017. Diffusion magnetic resonance imaging in cerebral small vessel disease. *Revue Neurologique* 173 (4), 201–210. <https://doi.org/10.1016/j.neuro.2017.03.005>.
- Magin R.L., Karaman, M.M., Hall, M.G., Zhu, W., Zhou, X.J. (2019). Capturing complexity of the diffusion-weighted MR signal decay. *Magn Reson Imaging*. Feb;56: 110-118. doi: 10.1016/j.mri.2018.09.034. Epub 2018 Oct 9. PMID: 30314665; PMID: PMC6348133.
- Mirzaalian, H., De Pierrefeu, A., Savadjiev, P., Pasternak, O., Bouix, S., Kubicki, M., Westin, C.F., Shenton, M.E., Rathi, Y., 2015. Harmonizing diffusion MRI data across multiple sites and scanners. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9349, 12–19. https://doi.org/10.1007/978-3-319-24553-9_2.
- Ning, L., Bonet-Carne, E., Grussu, F., Seppehrband, 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., ... Tax, C. M. W. (2020). Cross-scanner and cross-protocol multi-shell diffusion MRI data harmonization: Algorithms and results. *NeuroImage*, 221(October 2019). <https://doi.org/10.1016/j.neuroimage.2020.117128>.
- Palacios, E.M., Martin, A.J., Boss, M.A., Ezekiel, F., Chang, Y.S., Yuh, E.L., Vassar, M.J., Schnyer, D.M., MacDonald, C.L., Crawford, K.L., Irimia, A., Toga, A.W., Mukherjee, P., 2017. Toward precision and reproducibility of diffusion tensor imaging: A multicenter diffusion phantom and traveling volunteer study. *Am. J. Neuroradiol.* 38 (3), 537–545. <https://doi.org/10.3174/ajnr.A5025>.
- Reijmer, Y.D., Fotiadis, P., Martinez-Ramirez, S., Salat, D.H., Schultz, A., Shoamanesh, A., Ayres, A.M., Vashkevich, A., Rosas, D., Schwab, K., Leemans, A., Biessels, G.J., Rosand, J., Johnson, K.A., Viswanathan, A., Gurol, M.E., Greenberg, S. M., 2015. Structural network alterations and neurological dysfunction in cerebral amyloid angiopathy. *Brain* 138 (1), 179–188. <https://doi.org/10.1093/brain/awu316>.
- Reijmer, Y.D., Leemans, A., Brundel, M., Kappelle, L.J., Biessels, G.J., 2013. Disruption of the cerebral white matter network is related to slowing of information processing speed in patients with type 2 diabetes. *Diabetes* 62 (6), 2112–2115. <https://doi.org/10.2337/db12-1644>.
- Rydhög, A.S., Szczepankiewicz, F., Wirestam, R., Ahlgren, A., Westin, C.F., Knutsson, L., & Pasternak, O. (2017). Separating blood and water: Perfusion and free water elimination from diffusion MRI in the human brain. *NeuroImage*, 156(December 2016), 423–434. <https://doi.org/10.1016/j.neuroimage.2017.04.023>.
- 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 (4), 1487–1505. <https://doi.org/10.1016/j.neuroimage.2006.02.024>.
- Steven, A.J., Zhuo, J., Melhem, E.R., 2014. Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain. *AJR. Am. J. Roentgenol.* 202 (1), W26–W33.
- Takao, H., Hayashi, N., Kabasawa, H., Ohtomo, K., 2012. Effect of scanner in longitudinal diffusion tensor imaging studies. *Hum. Brain Mapp.* 33 (2), 466–477. <https://doi.org/10.1002/hbm.v33.210.1002/hbm.21225>.
- 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(August 2018), 285–299. <https://doi.org/10.1016/j.neuroimage.2019.01.077>.
- Teipel, S.J., Reuter, S., Stieltjes, B., Acosta-Cabrero, J., Ernemann, U., Fellgiebel, A., Filippi, M., Frisoni, G., Hentschel, F., Jessen, F., Klöppel, S., Meindl, T., Pouwels, P.J. W., Hauenstein, K.H., Hampel, H., 2011. Multicenter stability of diffusion tensor imaging measures: A European clinical and physical phantom study. *Psychiatry Research - Neuroimaging* 194 (3), 363–371. <https://doi.org/10.1016/j.psychres.2011.05.012>.
- Timmermans, C., Smeets, D., Verheyden, J., Terzopoulos, V., Anania, V., Parizel, P.M., Maas, A., 2019. Potential of a statistical approach for the standardization of multicenter diffusion tensor data: A phantom study. *J. Magn. Reson. Imaging* 49 (4), 955–965. <https://doi.org/10.1002/jmri.v49.410.1002/jmri.26333>.
- Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23 (3), 1176–1185. <https://doi.org/10.1016/j.neuroimage.2004.07.037>.
- Tuladhar, A.M., Van Norden, A.G.W., De Laat, K.F., Zwiers, M.P., Van Dijk, E.J., Norris, D.G., De Leeuw, F.E., 2015. White matter integrity in small vessel disease is related to cognition. *NeuroImage: Clinical* 7, 518–524. <https://doi.org/10.1016/j.nicl.2015.02.003>.
- Van Horn, J.D., Toga, A.W., 2014. Human neuroimaging as a “Big Data” science. *Brain Imaging and Behavior* 8 (2), 323–331. <https://doi.org/10.1007/s11682-013-9255-y>.
- van Leijsen, E.M.C., van Uden, I.W.M., Ghafoorian, M., Bergkamp, M.I., Lohner, V., Kooijmans, E.C.M., van der Holst, H.M., Tuladhar, A.M., Norris, D.G., van Dijk, E.J., Rutten-Jacobs, L.C.A., Platel, B., Klijn, C.J.M., de Leeuw, F.-E., 2017. Nonlinear temporal dynamics of cerebral small vessel disease. *Neurology* 89 (15), 1569–1577. <https://doi.org/10.1212/WNL.0000000000004490>.
- Van Norden, A.G.W., De Laat, K.F., Van Dijk, E.J., Van Uden, I.W.M., Van Oudheusden, L.J.B., Gons, R.A.R., Norris, D.G., Zwiers, M.P., De Leeuw, F.E., 2012. Diffusion tensor imaging and cognition in cerebral small vessel disease. *The RUN DMC study. Biochimica et Biophysica Acta - Molecular Basis of Disease* 1822 (3), 401–407. <https://doi.org/10.1016/j.bbadis.2011.04.008>.
- 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. <https://doi.org/10.1016/j.neuroimage.2013.05.028>.
- Vollmar, C., O’Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V., Duncan, J.S., Richardson, M.P., Koeppe, M.J., 2010. Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *NeuroImage* 51 (4), 1384–1394. <https://doi.org/10.1016/j.neuroimage.2010.03.046>.
- Vos, S.B., Jones, D.K., Jeurissen, B., Viergever, M.A., Leemans, A., 2012. The influence of complex white matter architecture on the mean diffusivity in diffusion tensor MRI of the human brain. *NeuroImage* 59 (3), 2208–2216. <https://doi.org/10.1016/j.neuroimage.2011.09.086>.
- Vos, S.B., Jones, D.K., Viergever, M.A., Leemans, A., 2011. Partial volume effect as a hidden covariate in DTI analyses. *NeuroImage* 55 (4), 1566–1576. <https://doi.org/10.1016/j.neuroimage.2011.01.048>.
- Vos, Sjoerd B., Tax, Chantal M.W., Luijten, Peter R., Ourselin, Sebastien, Leemans, Alexander, Froeling, Martijn, et al., 2016. The importance of correcting for signal drift in diffusion MRI. *Magnetic Resonance in Medicine* 77 (1), 285–299. <https://doi.org/10.1002/mrm.26124>.
- Wiegertjes, K., Ter Telgte, A., Oliveira, P.B., Van Leijsen, E.M.C., Bergkamp, M.I., Van Uden, I.W.M., Ghafoorian, M., Van Der Holst, H.M., Norris, D.G., Platel, B., Klijn, C. J.M., Tuladhar, A.M., De Leeuw, F.E., 2019. The role of small diffusion-weighted imaging lesions in cerebral small vessel disease. *Neurology* 93 (17), E1627–E1634. <https://doi.org/10.1212/WNL.0000000000008364>.
- Wollenweber, F.A., Hanecker, P., Bayer-Karpinska, A., Malik, R., Bänzner, H., Moreton, F., Muir, K.W., Müller, S., Giese, A., Opherk, C., Dichgans, M., Haffner, C., Düring, M., 2015. Cysteine-sparing CADASIL mutations in NOTCH3 show proaggregatory properties in vitro. *Stroke* 46 (3), 786–792. <https://doi.org/10.1161/STROKEAHA.114.007472>.