frontiers in Neuroscience

REVIEW
published: 06 May 2020
doi: 10.3389/fnins.2020.00396

# Harmonization of Brain Diffusion MRI: Concepts and Methods

*Maíra Siqueira Pinto[1,2*†], Roberto Paolella[2,3†], Thibo Billiet[3], Pieter Van Dyck[1], Pieter-Jan Guns[4], Ben Jeurissen[2], Annemie Ribbens[3], Arnold J. den Dekker[2] and Jan Sijbers[2]*

[1] Department of Radiology, Antwerp University Hospital, University of Antwerp, Antwerp, Belgium, [2] imec–Vision Lab, University of Antwerp, Antwerp, Belgium, [3] Icometrix, Leuven, Belgium, [4] Physiopharmacology, University of Antwerp, Antwerp, Belgium

MRI diffusion data suffers from significant inter- and intra-site variability, which hinders multi-site and/or longitudinal diffusion studies. This variability may arise from a range of factors, such as hardware, reconstruction algorithms and acquisition settings. To allow a reliable comparison and joint analysis of diffusion data across sites and over time, there is a clear need for robust data harmonization methods. This review article provides a comprehensive overview of diffusion data harmonization concepts and methods, and their limitations. Overall, the methods for the harmonization of multi-site diffusion images can be categorized in two main groups: diffusion parametric map harmonization (DPMH) and diffusion weighted image harmonization (DWIH). Whereas DPMH harmonizes the diffusion parametric maps (e.g., FA, MD, and MK), DWIH harmonizes the diffusion-weighted images. Defining a gold standard harmonization technique for dMRI data is still an ongoing challenge. Nevertheless, in this paper we provide two classification tools, namely a feature table and a flowchart, which aim to guide the readers in selecting an appropriate harmonization method for their study.

**Keywords: harmonization, normalization, diffusion MRI, multi-site, inter-scanner, review**

OPEN ACCESS

***Edited by:***
*Alard Roebroeck, Maastricht University, Netherlands*

***Reviewed by:***
*Daniel Güllmar, Friedrich-Schiller University of Jena, Germany*
*Suyash P. Awate, Indian Institute of Technology Bombay, India*

****Correspondence:***
*Maíra Siqueira Pinto*
*mairaspinto@gmail.com*

*† These authors have contributed equally to this work*

***Specialty section:***
*This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience*

***Received:*** *18 July 2019*
***Accepted:*** *30 March 2020*
***Published:*** *06 May 2020*

***Citation:***
*Pinto MS, Paolella R, Billiet T, Van Dyck P, Guns P-J, Jeurissen B, Ribbens A, den Dekker AJ and Sijbers J (2020) Harmonization of Brain Diffusion MRI: Concepts and Methods. Front. Neurosci. 14:396. doi: 10.3389/fnins.2020.00396*

## INTRODUCTION

Diffusion-weighted magnetic resonance imaging (dMRI) is an MRI technique in which the image contrast is related to the diffusion of water molecules inside tissues. dMRI has brought great innovation to neuroimaging analysis, since it enables non-invasive probing of brain microstructure. Nevertheless, many studies using diffusion data rely on small sample sizes, leading to poor reproducibility of results. Fortunately, research is evolving toward large multicenter studies with the aim of increasing statistical power. However, the success of a joint analysis is highly dependent on the comparability of the multi-site data.

Diffusion data of the same subject obtained at different sites and/or acquired at different time points can be different due to local and/or temporal scanner characteristics resulting in a high inter- and intra-scanner variability (Vollmar et al., 2010; Grech-Sollars et al., 2015; Nencka et al., 2017). These variabilities may arise from a range of factors, such as hardware (scanner manufacturer, field strength, transmitter/receiver coils, magnetic field inhomogeneities, etc.), reconstruction algorithms (SENSE, GRAPPA, etc.), acquisition parameters (voxel size, number of gradient directions, echo time, repetition time, etc.), and image quality [signal to noise ratio (SNR), etc.] (Alexander et al., 2006; Ni et al., 2006; Jones, 2010; Vollmar et al., 2010). All these factors affect

the final diffusion signal intensity and consequently the diffusion metrics, preventing reliable multi-site and/or longitudinal diffusion studies (Pfefferbaum et al., 2003; Vollmar et al., 2010; Mirzaalian et al., 2018).

In literature, many conflicting inferences have been reported between studies, in which findings based on small distinct cohorts are used to generalize conclusions for an entire population, without considering intra- and inter-site differences (Button et al., 2013; Kelly et al., 2018; Smith and Nichols, 2018). To determine the site effects on diffusion data, a number of studies examined diffusion phantom data to detect scanner related variabilities (Teipel et al., 2011; Zhu et al., 2011; Walker et al., 2013; Pullens et al., 2017; Timmermans et al., 2019). Up to 7% of inter-site variability in diffusion metrics was demonstrated in phantoms (Teipel et al., 2011; Palacios et al., 2017). However, using parameters obtained from phantom data to correct human data is not advised due to the structural complexity of human biological tissue (Karayumak et al., 2019).

Previous research has established that inter-site variability is non-uniform across the white matter of the human brain, with a variability up to 5% in diffusion metrics of major brain tracts (Vollmar et al., 2010; Grech-Sollars et al., 2015; Nencka et al., 2017). Recently, investigators have examined the reproducibility of multi-shell diffusion images in a multi-site study involving traveling subjects (Tong et al., 2019). A 7.7% median inter-center coefficient of variation was estimated for the track density maps in whole white matter among the subjects. These inter-site variabilities in diffusion metrics are similar to the changes due to pathologies. For example, in the work of Kumar et al. (2009), it was shown that the variability in diffusion metrics in the corpus callosum between controls, mild Traumatic Brain Injury (TBI) and moderate TBI patients, are of the same order as intra-scanner changes. Furthermore, a quantitative study by Mahoney et al. reported longitudinal changes in diffusion metrics in dementia patients compared to controls in the same order of the site variabilities (Mahoney et al., 2015). From these findings, we can infer that it is crucial to reduce the variability across multi-center diffusion data.

Inter-site variability can be reduced by acquiring data with scanners from the same manufacturer at each site and using similar acquisition parameters (Vollmar et al., 2010; Fox et al., 2012; Cannon et al., 2014). However, diffusion parameters of subjects scanned using the same acquisition protocol may still differ significantly across sites (Nyholm et al., 2013; Jovicich et al., 2014; Mirzaalian et al., 2015). These differences may come from several sources, such as sensitivity of head coils, imaging gradient non-linearities, magnetic field inhomogeneities and other scanner related factors. Hence, there is a substantial need for robust harmonization techniques (Jenkins et al., 2016; Jovicich et al., 2019). The overall concept of harmonization methods is to apply statistical or mathematical concepts to reduce unwanted site variability while maintaining the biological content. In the last decade a multitude of harmonization methods have been developed.

For this review, we have categorized the brain dMRI methods in two main groups depending on the data-format used as input for harmonization. The first category uses calculated diffusion (para)metric maps, such as Fractional Anisotropy (FA), Mean, Axial and Radial Diffusivity (MD, AD, and RD, respectively), Kurtosis Anisotropy (KA), Mean, Axial, and Radial Kurtosis (MK, AK, and RK, respectively), as input (i.e., diffusion parametric map harmonization; DPMH). While the second category uses diffusion weighted images (DWI) as input (i.e., diffusion weighted image harmonization; DWIH).

To the authors' knowledge, no previous study has provided an extensive report of diffusion harmonization methods. In this review paper, a comprehensive overview of those methods is presented, including an investigative analysis of their strengths and weaknesses. DPMH and DWIH methods reported since 2009 are described. This paper is organized as follows. Section "Literature Search" describes the search mechanism used for selecting the literature on brain diffusion data harmonization. In Section "Requirements for Harmonization," the requirements for harmonization are specified. Sections "Diffusion Parametric Map Harmonization Methods" and "Diffusion Weighted Image Harmonization Methods" depict the DPMH and DWIH harmonization methods reported in the literature. Section "Discussion" then presents an overview of the main characteristics of the methods and a guideline that helps the user to select an adequate harmonization method for her/his data. Finally, in Section "Conclusion" conclusions are drawn.

## LITERATURE SEARCH

Two authors (MSP and RP) independently performed a literature search across two databases (PubMed and Google Scholar) using combinations of the following search terms: "harmonization," "harmonisation," "normalization," "normalisation," "multi-site," "multi-center," "inter-site," "intra-scanner," "diffusion," "MRI," "DTI," "meta-analysis," "covariates," "spherical harmonics," "deep learning." Besides the usual search engines, additional important papers were selected by checking the reference lists of identified relevant publications on data harmonization. After removing the duplicates, all identified articles were screened by title and abstract. Studies were included if they described diffusion harmonization methods and concepts.

## REQUIREMENTS FOR HARMONIZATION

For the majority of dMRI harmonization procedures, co-registration is of crucial importance. Co-registration of diffusion images aims to find spatial transformations to map different images to a common reference space, allowing direct comparison of various image properties. Prior to harmonization, a voxel-by-voxel correspondence between multiple diffusion volumes is needed, in order to minimize errors in subsequent calculations. In particular, voxel-wise DPMH and DWIH approaches require all subjects to be in the same space in order to extract common features that are

site-related rather than anatomically specific. The common space can be a study-specific template or a standard brain atlas template, as for example, the ICBM152 template of the Montreal Neurological Institute (MNI) space[1]. Many tools are available for registering diffusion images, such as Advanced Normalization Tools (ANTs; Avants et al., 2011), FMRIB Software Library (FSL; Jenkinson et al., 2012), and elastix (Klein et al., 2010; Shamonin et al., 2014).

Additionally, a dataset with a balanced number of subjects per site is advised for robust harmonization. Many DPMH and DWIH methods use these subjects to efficiently learn a set of so-called mapping parameters used to characterize the differences between the images across scanners. Additionally, an important requirement, especially for DWIH methods, is the availability of training data, i.e., matched subjects across sites for obtaining the mapping parameters between sites. Age, gender, handedness, and socio-economic status need to be matched among the subjects to remove the statistical differences at group level. Moreover, for some machine learning techniques, there is a need for DWI data of individual subjects that are scanned at different sites, within a small interval of time, to train a network to recognize site-related underlying inter-scanner/inter-site differences in the characteristics of the images to harmonize.

Overall, for all the methods, it is highly recommended to use a balanced dataset and to co-register the diffusion images or maps to a common template. The recommendations are to assure that statistical differences are only due to hardware, software and protocol differences, and ensure spatial compatibility intra- and inter-subjects during the harmonization procedure. Furthermore, each method has its own specifications and limitations that are described in the following sections.

# DIFFUSION PARAMETRIC MAP HARMONIZATION METHODS

Diffusion parametric map harmonization methods perform particular transformations on the diffusion parametric maps that enable data pooling and reduction of unwanted intra- and inter-site variability. For a joint analysis of multi-site diffusion metric maps that have been estimated using a given diffusion model [e.g., diffusion tensor imaging (DTI), diffusion kurtosis imaging (DKI), neurite orientation dispersion and density imaging (NODDI), etc.], statistical or mathematical DPMH methods can be applied. The purpose of these methods is to perform joint statistical analysis on multi-site data. It can be performed in two ways: (1) without modifying the original diffusion parametric maps (see Subsection "Modeling Inter-Site Variability Within the Statistical Analysis"); (2) by modifying the parametric maps with *a posteriori* analysis (see Subsection "Harmonizing the Parametric Maps Based on Regression of Covariates"). DPMH methods allow to pool DWI parametric maps obtained from different diffusion acquisition schemes (diffusion directions, *b*-values, repetition time, echo time, etc.). The DPMH methods described below are meta-analysis, mega-analysis, and regression of covariates.

[1]http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009

## Modeling Inter-Site Variability Within the Statistical Analysis

### Meta-Analysis

Meta-analysis is a popular statistical analysis technique in biomedical research that combines results of independent multi-site and/or longitudinal studies. The general concept is to perform a group-wise statistical analysis separately for each site, followed by a weighted combination of effect size over the different studies to strengthen conclusions about the research question (Zhu et al., 2019). Meta-analysis is useful to pool retrospective data with sample sizes that are too small to draw valid conclusions independently (Petitti, 1994).

**Figure 1** presents an example of meta-analysis in which statistical inferences are obtained independently per site from the FA maps of different groups of subjects. As a first step, an intra-site statistical analysis is performed. The resulting statistical scores (e.g., *z*-score) of the metric of interest (e.g., FA) can then be weighted by each site's sample size or with respect to an estimate of precision, such as effect size (Salimi-Khorshidi et al., 2009), to obtain the final statistical score. In contrast to this approach, the overall statistical score can also be obtained by modeling site as a random effect (Worsley, 2002; Beckmann et al., 2003; Woolrich et al., 2004). For example, in the work of Teipel et al. (2012), meta-analysis was used to investigate FA and MD differences between dementia patients and controls in a multi-site study, taking scanner effects into account. Voxel-based *t*-statistics were converted to *z*-scores after which a variance component analysis was applied, effectively reducing effects of site (random effect), age and gender (fixed effects).

One of the main advantages of meta-analysis is the possibility to pool data from small/underpowered studies to derive robust conclusions. It is also the only way to pool studies for which only aggregated data are reported (e.g., group difference statistics or the mean FA per region of interest) and for which the whole brain images are not available. However, one drawback is that if the statistics performed in the individual studies are biased by study size, the population estimate will be also affected. Another disadvantage is that the statistical analysis should first be performed separately for each diffusion metric of interest.

### Mega-Analysis

In contrast to meta-analysis, mega-analysis refers to a technique of summarizing the statistics from the individual subjects of all sites to jointly evaluate population group differences (Jahanshad et al., 2013; Zhu et al., 2019). As depicted in **Figure 1**, in mega-analysis group-difference statistics are not calculated for each site separately. Instead, group differences are identified by a site-weighted combination of the statistical scores from all individuals jointly.

When the individual diffusion data (e.g., FA) is available per subject, the measures can be pooled to calculate the effect size across the entire group in a mega-analysis. To take into account the variability due to site differences, the site effect can be modeled using a mixed linear model statistical approach (or another statistical method to analyze the dataset), just as in meta-analysis.

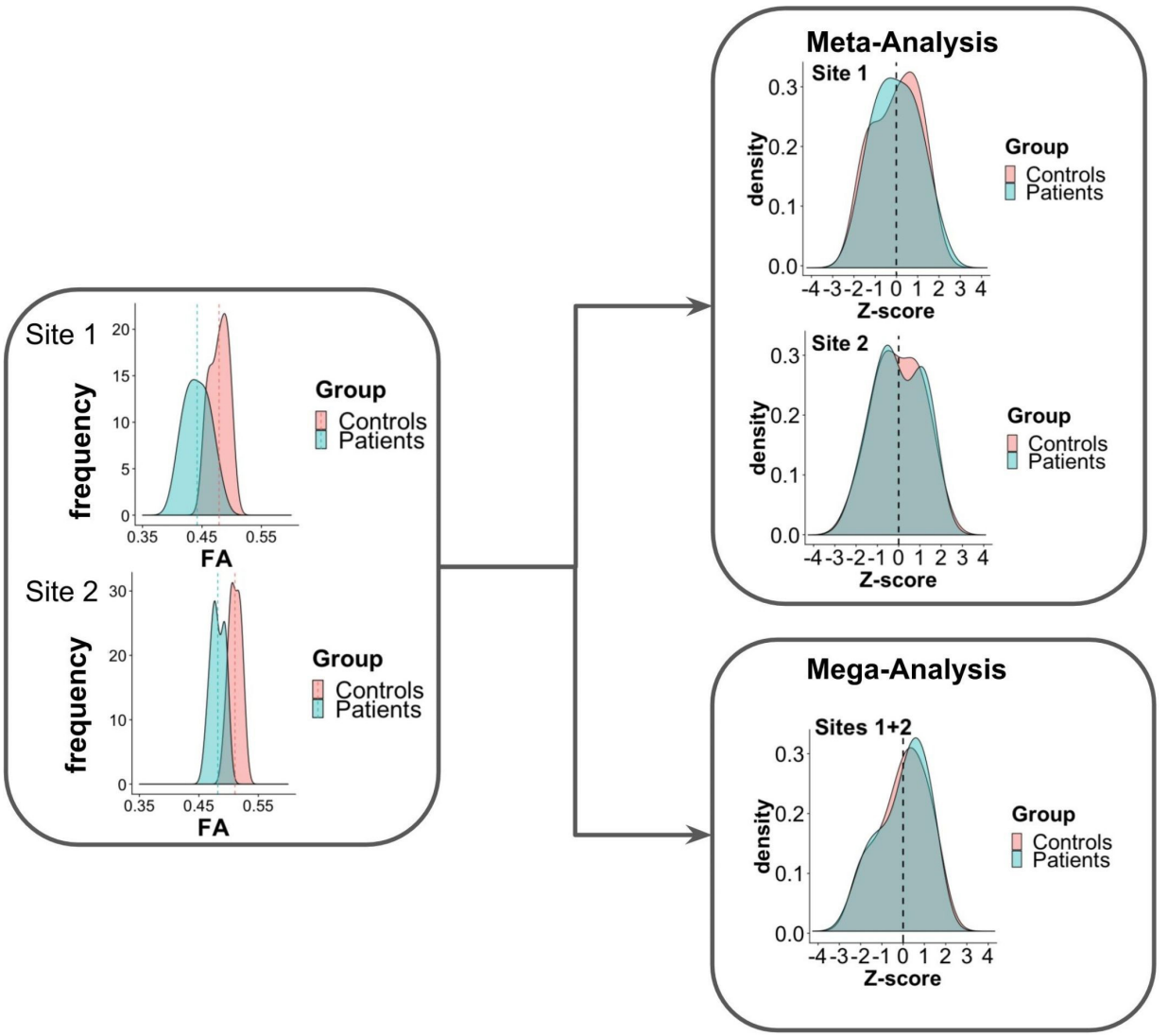


**FIGURE 1 |** Scheme of meta- and mega-analysis. FA measures from sites 1 and 2, for two groups of subjects: controls and patients. The FA frequency for each group is estimated for each of the sites. Meta-analysis performs the statistical evaluation between groups for each site separately, followed by a weighted combination of its statistical results, while in mega-analysis a weighted statistical evaluation is performed for all sites jointly.

While not directly harmonizing the imaging data itself, mega-analysis allows a joint analysis of two (or more) datasets to evaluate a common characteristic in the population (Jahanshad et al., 2013; Kochunov et al., 2014; Zhu et al., 2019). Some limitations in this approach are that the size of the cohort may not be sufficient to capture the variance of the entire population, pre-processing steps could be very different for each site (if the FA maps are computed independently), and the statistical analysis has to be performed separately for each variable (e.g., FA, MD, AD, and RD).

Meta-and mega-analysis have successfully been adopted in the field of neuroimaging by the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium (Jahanshad et al., 2013; Kochunov et al., 2014). The general concept of the harmonization method proposed by the ENIGMA-DTI group is that each site preprocesses the diffusion metric maps (e.g., FA) separately. The statistical scores are harmonized using meta- or mega-analysis, to improve data comparability and robustness. Findings of the ENIGMA-DTI group indicate that results obtained by meta- and mega-analysis may differ, in favor of the latter. In multi-center studies with a moderate amount of variation between cohorts, a mega-analysis statistical framework appears to be the better approach to investigate structural neuroimaging data, showing greater stability and higher power for jointly analyzing the data (Kochunov et al., 2014). Nonetheless, when the individual diffusion metric maps are not available, meta-analysis could serve as a valuable alternative. However, meta-analysis should be performed carefully and one should take into account cohort trends (Kochunov et al., 2014).

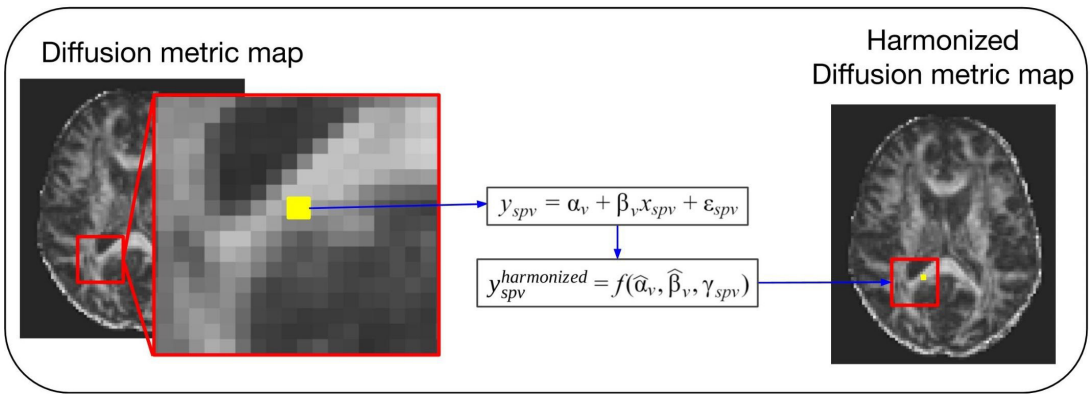


**FIGURE 2 |** General scheme of a voxel-wise regression of covariates harmonization approach. For these methods the voxel intensity of the diffusion metric maps ($y_{spv}$, the intensity for a specific site *s*, subject *p* and voxel *v*) is modeled as a combination of a voxel-wise intercept ($\alpha_v$), a voxel-wise slope ($\beta_v$) multiplied by a model-specific dependent variable ($x_{spv}$), and an error component ($\varepsilon_{spv}$). Each of the regression of covariates approaches will have a different model and dependent variable to describe the biological and site-related effects on the diffusion metric intensities. Next, the estimated coefficients are used to compute the new harmonized diffusion intensity values ($y_{spv}^{harmonized}$).

## Harmonizing the Parametric Maps Based on Regression of Covariates

Covariates, also known as explanatory variables, are variables that may affect the estimate of the diffusion metric under study. These covariates can be variables of clinical interest or unwanted confounding variables, such as MR hardware (e.g., scanner manufacturer, field strength, and coils), software, acquisition parameters (e.g., echo time, repetition time, *b*-value, and gradient directions) or image quality. One way to handle unwanted variability due to confounding factors is the use of regression models (Pourhoseingholi et al., 2012). This approach is illustrated in **Figure 2**. After fitting a regression model to the diffusion values, adjusted values can be derived that no longer contain the effect of the covariates. The use of the regression of covariates harmonization approach to correct for variability in software and hardware has been reported extensively in the literature (Forsyth et al., 2014; Venkatraman et al., 2015; Fortin et al., 2016, 2017; Pohl et al., 2016; Timmermans et al., 2019). Regression of covariates methods can be divided into two categories: global harmonization methods and voxel-wise harmonization methods. Both classes are described below.

The methods present different options for harmonizing diffusion metric maps (e.g., FA and MD maps). For briefness, we use the notation $y_{spv}^{method}$ to denote the diffusion metric measure *y* harmonized by a specific *method*, at site *s*, for subject *p* and voxel *v*.

### Global Harmonization

#### *Human-phantom based harmonization (HuP)*

A straightforward approach for data harmonization is to apply scanner-specific correction factors derived from human phantom data (i.e., a group of individuals scanned at multiple scanners/sites within a short period of time) (Pohl et al., 2016). One scanner type is defined as the reference (*R*) and the other as the target (*T*). The goal is to correct the diffusion metric maps of the target site. For this purpose, a correction factor (*F*) is calculated as the ratio of the mean value (across the human phantoms) of the diffusion metric in the reference and target, respectively: $F = \frac{\sum_p \bar{y}_p^R / N}{\sum_p \bar{y}_p^T / N}$, where $\bar{y}_p^R$ and $\bar{y}_p^T$ are the mean metric value across the white matter voxels for human phantom *p* at the reference and target site, respectively, and *N* is the number of human phantoms. Successively, once the scanner-specific correction factors are determined, metric maps *y* for subject *p* and voxel *v* scanned in the target scanner ($y_{spv}$) are scaled by the appropriate correction factor in order to obtain the HuP-harmonized diffusion metric maps: $y_{spv}^{HuP} = y_{spv} F$.

The main advantages of the correction factor are its simple derivation and the fact that it has been demonstrated to correct for differences that are likely attributable to the MR system manufacturer (Pohl et al., 2016). However, human phantom datasets from multiple sites are required. Moreover, a unique correction factor per scanner type only partially reduces the harmonization problem due to its intrinsic non-linearity, i.e., scanner type differences are not uniform but vary in a highly non-linear fashion across the brain (Karayumak et al., 2019).

#### *Hardware-phantom based harmonization (HaP)*

Timmermans et al. (2019) presented global multi-site harmonization models, using phantom data acquired at multiple centers in a longitudinal study. For this study, dedicated diffusion single-strand phantoms were developed by HQ Imaging (Heidelberg, Germany). The study aimed to build a comprehensive model for the variability of FA. Protocol-specific and site-specific effects were included in the models, considering hardware (scanner vendor and head coil), software, acquisition parameters (bandwidth, TE, and TR), image quality (signal-to-noise ratio and mean residual), as fixed predictor variables, and site as random predictor variable, taking into account that

fixed predictors relate to effects that are constant across all individuals, and random predictors relate to effects that vary across individuals.

Different models were proposed to describe the diffusion metric values $y_p$ of the phantoms $p$ considering the differences between acquisitions and were evaluated via the combination of the fixed and random predictors ($x_p$ and $z_p$, respectively): $y_p^{HaP} = \beta_0 + \beta_n x_p + b_{0p} + b_{np} z_p + \varepsilon_p$, where $\beta_0$ is the fixed intercept, $\beta_n$ the fixed effects slope, $b_{0p}$ the random intercept per phantom, $b_{np}$ the random slope per phantom, and $\varepsilon_p$ the error. In order to find the most comprehensive model for the diffusion metric data, many linear mixed effects models were evaluated by the Akaike information criterion (AIC). The selection of model parameters was based on three model categories: protocol-specific intercept, protocol-specific intercept with quality effects, and protocol-specific intercept with protocol-specific quality effects. Each model was further divided into submodels depending on the included variables. AIC is used to select which model best describes the variations in the metric intensities. The results showed that scanner manufacturer, SNR, head coil, bandwidth and TE are the covariates that best describe the sources of variability in the inter-site phantom data, and should be used to harmonize the diffusion metric maps of multi-center studies.

The use of hardware phantoms for harmonization has several advantages. Hardware phantoms can be scanned multiple times, for a longer time, and their images do not suffer from motion artifacts. The phantom content is controllable and remains stable over time. Duplicated phantoms can be easily obtained by several sites, obviating transport. The main drawback of hardware phantom based harmonization is that such phantoms do not fully represent the complexity of the human brain, and therefore have different, intra- and inter-scanner variabilities. Obviously, voxel-wise harmonization (cf., Section "Voxel-Wise Harmonization") of brain dMRI is not possible using phantom data.

### *Global scaling (GS)*

In the global scaling method presented by Fortin et al., 2017, a linear model is used to correct the site effect on the diffusion metric maps (Fortin et al., 2017). The estimated location ($\theta_{s,location}$) and scale ($\theta_{s,scale}$) model parameters, per site $s$, encapsule the variabilities in the diffusion metric maps due to site effects. They are estimated by fitting a linear regression model: $\bar{Y}_s = \theta_{s,location} + \theta_{s,scale}\,\bar{Y} + \varepsilon_s$, where $\bar{Y}_s$ is an $n_v \times 1$ vector containing the average diffusion metric intensity per voxel for the number of voxels $n_v$ computed over all subjects of site $s$, $\bar{Y}$is an $n_v$ x1 vector containing the average diffusion metric intensity per voxel for the number of voxels $n_v$ computed over all subjects of all sites together (considered a reference), and $\varepsilon_s$ is the residual error. From the estimated parameters, the harmonized diffusion metric maps are calculated as: $y_{spv}^{GS} = \frac{y_{spv} - \hat{\theta}_{s,location}}{\hat{\theta}_{s,scale}}$ .

The main advantage of global scaling is that it takes into account information from all sites. Some disadvantages are that the removal of site effects can also remove biological variability, and that it does not account for spatial heterogeneity of the site effects in the brain.

## Voxel-Wise Harmonization

### *Removal of Artificial Voxel Effect by Linear regression (RAVEL)*

The Removal of Artificial Voxel Effect by Linear regression (RAVEL) method (Fortin et al., 2016) uses voxels in the cerebrospinal fluid (CSF) voxels as control region. The CSF-voxels are used for harmonization because their diffusion metric intensities are unassociated with disease or other clinical factors and are theoretically only influenced by site-related variabilities. In this method, the voxel-wise intensity of the diffusion metric maps ($y_{spv}$) is described as a combination of four components: the average intensity in the sample ($\alpha 1^t$), the known clinical covariates of interest ($\beta X^t$), the unknown site-related factors ($\gamma Z^t$) and a residual ($R$): $y_{spv} = \alpha 1^t + \beta X^t + \gamma Z^t + R$. Where the symbol $t$ indicates the transpose operation, $y_{spv}$ is the $v \times p$ matrix containing the registered and normalized voxel intensities for $v$ voxels and $p$ subjects, $\alpha 1^t$ is a $v \times 1$ vector containing the average voxel intensity per site, $X$ is a $p \times k$ matrix containing for each subject $p$ the correspondent biological covariates $k$, $\beta$ is the coefficient matrix associated with $X$, $Z$ is a $p \times m$ matrix containing for each subject $p$ the associated $m$ unwanted coefficient factors and $\gamma$ is the coefficient matrix associated with $Z$.

The CSF voxels are used to estimate the unknown/unwanted factors ($Z^t$) by assuming that $\alpha$ and $\beta$ are null for the CSF since there is no association between control voxels and clinical features. Thus, the CSF diffusion intensities ($y_{spv}^{CSF}$) are described as: $y_{spv}^{CSF} = \gamma_{CSF} Z^t + R_{CSF}$. Singular value decomposition is used to obtain the first latent factors ($w_{1sp}$) from the CSF voxels, representing the site-related variability common to all voxels. Next, the voxel-wise RAVEL coefficients ($\psi_v$) are estimated fitting the linear regression model to the voxel-wise diffusion intensities ($y_{spv}$) and the first latent factors ($w_{1sp}$): $y_{spv} = \alpha_v + \psi_v w_{1sp} + \varepsilon_{spv}$, where $\varepsilon_{spv}$ is the residual error. Lastly, the RAVEL-harmonized diffusion metric map intensities are computed: $y_{spv}^{RAVEL} = y_{spv} - \hat{\psi}_v w_{1sp}$ .

An advantage of the RAVEL method is that it is a voxel-wise harmonization method that uses intra-subject information that is not affected by disease (CSF control region) for improving comparability between subjects. However, if these control regions do not carry the information about the inter-site variability and/or are related to the parameter of interest, then the correction may remove relevant biological information, becoming a disadvantage to use this method in such cases.

### *Surrogate Variable Analysis (SVA)*

Surrogate Variable Analysis (SVA) identifies and estimates unknown, unmodeled or unwanted sources of variation from the data (Leek et al., 2012; Fortin et al., 2017). The so-called batch effects can be defined as measurements of unwanted variability that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study (Leek et al., 2010). In the context of multi-site harmonization, SVA is particularly useful when it is not known which datasets belong to which site. Through singular value decomposition, the data is decomposed into a set of $m$ surrogate variables ($z_1$, ..., $z_m$). Variables with the largest variance, and which are not covarying

with *a priori* defined factors of interest such as age, gender or diagnosis, are then regressed out of the data. The voxel-wise SVA coefficients ($\Phi_{mv}$) are estimated by fitting the surrogate variables ($z_{msv}$, for surrogate variable $m$, site $s$ and voxel $v$) and the original diffusion metric intensities ($y_{spv}$, for site $s$, subject $p$ and voxel $v$) to the linear regression model: $y_{spv} = \alpha_v + \sum_{n=1}^{m} \Phi_{nv} z_{nsp} + \varepsilon_{spv}$, where $\alpha_v$ is the voxel-wise overall measure of the diffusion metric and $\varepsilon_{spv}$ is the residual error. Next, the SVA-harmonized diffusion metric map intensities ($y_{spv}^{SVA}$) are computed as: $y_{spv}^{SVA} = y_{spv} - \sum_{n=1}^{m} \hat{\Phi}_{nv} z_{nsp}$.

Surrogate variable analysis is implemented in the SVA package for R, and is applicable voxel-wise (Leek et al., 2012). A strong point is that it estimates all common sources of latent variation, without needing to know their exact origin (e.g., site). Nonetheless, if this inherent variation is related to biological variability (e.g., patients in site A, controls in site B) then SVA is not appropriate.

***Combined association test (ComBat)***

The combined association test (ComBat) uses regression of covariates for data harmonization (Fortin et al., 2017). It started as a batch effect correction tool (similar to SVA) used in genomics, in which the batch effect is known (Johnson et al., 2007). It is a powerful and fast alternative for SVA in cases where site is an *a priori* known factor.

ComBat describes the non-harmonized diffusion metric in each voxel ($y_{spv}$, for site $s$, subjects $p$ and voxel $v$) by an adjustment model that consists of the following terms: an overall measure of the diffusion metric ($\alpha_v$), the product of a design matrix ($X_{sp}$) containing the covariates of interest (e.g., gender and age) and the vector of corresponding regression coefficients ($\beta_v$), a term representing the so-called additive site effects ($\gamma_{sv}$) and, finally, the product of a normally distributed error term ($\varepsilon_{spv}$) and a factor representing the so-called multiplicative site effects ($\delta_{sv}$): $y_{spv} = \alpha_v + X_{sp}\beta_v + \gamma_{sv} + \delta_{sv}\varepsilon_{spv}$. The site-specific parameters of the adjustment model are assumed to have parametric prior distributions, being a normal distribution for the additive factor ($\gamma_{sv}$) and an inverse gamma distribution for the multiplicative factor ($\delta_{sv}$). The parametric distributions are estimated from the data, using an empirical Bayes framework to decrease the variance of the site effects. It assumes that all voxels share a common distribution, and are used to infer the properties of the site-effects. Subsequently, ComBat-harmonized diffusion parameter maps are created based on the estimated additive and multiplicative factors ($\gamma_{sv}^{*}$ and $\delta_{sv}^{*}$, respectively): $y_{spv}^{ComBat} = \frac{y_{spv} - \hat{\alpha}_v - X_{sp}\hat{\beta}_v - \gamma_{sv}^{*}}{\delta_{sv}^{*}} + \hat{\alpha}_v + X_{sp}\hat{\beta}_v$.

It was reported that the ComBat harmonization method preserves between-subject biological information (Fortin et al., 2017). However, a limitation of this method is that the optimization procedure assumes the site effect parameters to follow a particular parametric prior distribution (Gaussian and Inverse-gamma), which might not generalize to all scenarios or measures. Moreover, it is not clear how non-linearities in the signal due to site effects propagate through the preprocessing techniques, as well as model fitting procedures.

# DIFFUSION WEIGHTED IMAGE HARMONIZATION METHODS

Diffusion parametric map harmonization methods for data pooling and joint analysis, meta- and mega-analysis and regression of covariates, have been reported extensively in the literature. Nonetheless, the harmonization of diffusion metric maps has several drawbacks, as described in section 4 for each of the methods. Recall that one of the main drawbacks is the lack of knowledge on how the scanner-specific non-linearities propagate in the diffusion model fit, possibly affecting the harmonization procedure of the diffusion metric maps. Recently, the use of the dMRI intensity signal has been proposed to perform model-free harmonization approaches. These methods are categorized as DWIH (Mirzaalian et al., 2015; Koppers et al., 2018; Huynh et al., 2019; Karayumak et al., 2019; Tax et al., 2019). DWIH methods rely on mapping the DWI images to a reference space. An overview of these DWIH approaches is given below. The methods described are the rotation invariant spherical harmonics method, machine learning algorithms, and the method of moments.

## Rotation Invariant Spherical Harmonics (RISH)

The use of rotation invariant spherical harmonics (RISH) for dMRI signal harmonization has been first proposed by Mirzaalian et al. (2015) and several improvements to this method have been presented since then (Mirzaalian et al., 2016, 2018; Karayumak et al., 2019).

The core idea of the RISH method is to map the diffusion weighted imaging (DWI) data from a target ($T$) site to a reference ($R$) site. The voxel-wise DWI signal intensity $S = [s_1, \ldots, s_g]^t$, along $g$ unique directions, can be compactly represented in a spherical harmonics (SH) basis: $S \approx \sum_i \sum_j C_{ij} Y_{ij}$, composed by SH basis functions ($Y_{ij}$) and their corresponding coefficients ($C_{ij}$) of order $i$ and degree $j$, with $j$ = 1,2,...,2i + 1. The RISH features, per harmonic order, are extracted from the estimated SH coefficients as: $RISH_i = ||C_i^2|| = \sum_{j=1}^{2i+1} (C_{ij}^2)$.

The harmonization procedure, which is illustrated in **Figure 3**, consists of two parts: (1) learning scale maps between sites from training data and (2) applying the learned scale maps to harmonize all DWI of the target site. The learning part is performed using training data that is a subset of subjects that are matched by age and gender for both sites. From the DWI, the RISH features are calculated and used to create a multivariate template, per $b$-value shell. In template space, the voxel-wise expected value per site $s$ and per harmonic order $i$ [$E_i^s(v)$] of RISH features is calculated as the sample mean over the number of training subjects ($N_s$): $E_i^s(v) \approx \sum_{p=1}^{N_s} RISH_i^s(v,p)/N_s$, where $s$ represents the site, $v$ the voxel location in template space and $p$ the training subject. Then, voxel-wise scale maps ($\Phi_i$) are computed for each harmonic order i: $\Phi_i(v,R,T) = \sqrt{\frac{E_i^R(v)}{E_i^T(v)+\varepsilon}}$. Next, in the application part, the scale maps are used to calculate the harmonized SH coefficients of the target data per harmonic

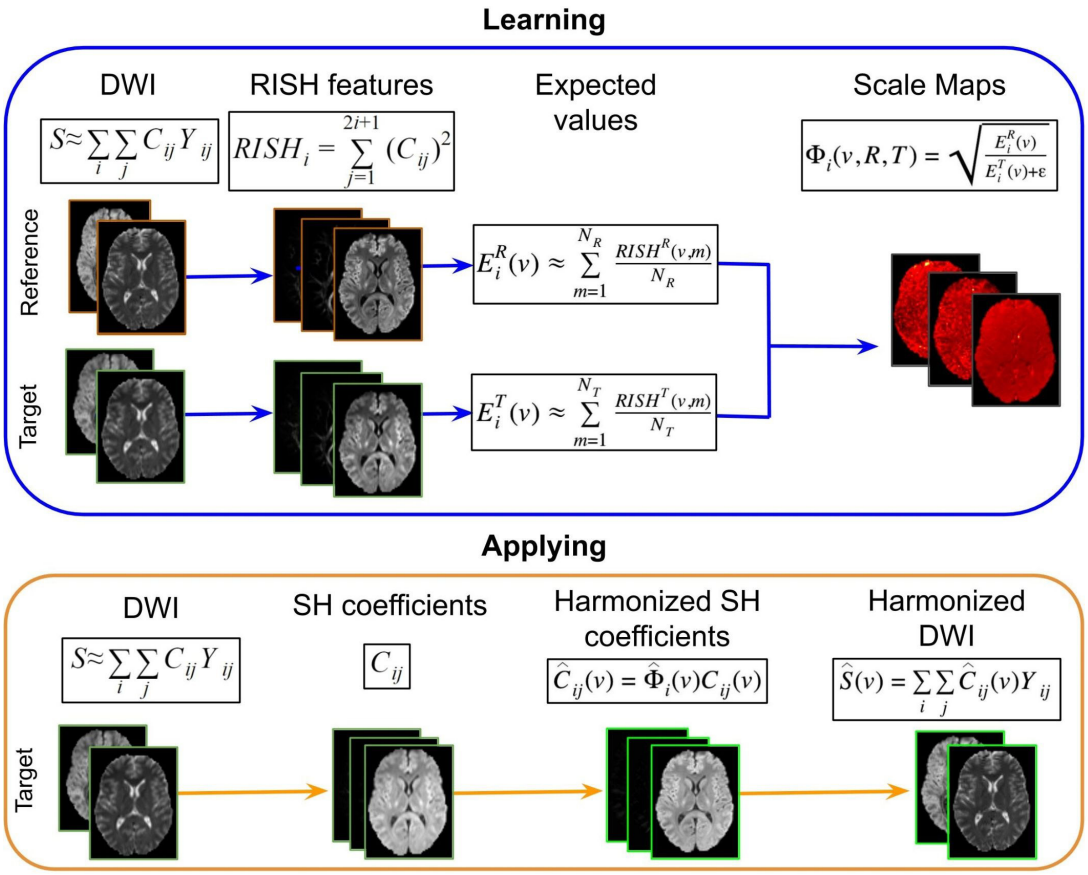


**FIGURE 3 |** Representation of the RISH harmonization approach. Consider the purpose of modifying the DWI acquired in a target site, to correspond to the DWI acquired in the reference site. In the learning part using matched subjects, the RISH features are computed in native space from the DWI for the two data sets separately: reference (*R*) and target (*T*) sites. Then RISH features are transformed to a common space, the expected values are calculated per site *s* and per harmonic order *i* ($E_i^s$), after which the scale maps are calculated ($\Phi_i$). The scale maps, which are computed for each harmonic order *i*, represent the transformation of the RISH features from target to reference site. Next, in the application step, the SH coefficients from the target site are calculated, the scale maps are warped into native space and applied to the SH coefficients, creating harmonized SH coefficients in native space. Those are transformed back to the signal intensity domain, obtaining the harmonized DWI. Thus, harmonized DWI from the target site can be jointly analyzed with the ones from the reference site.

order: $\hat{C}_{ij}(v) = \hat{\Phi}_i(v)C_{ij}(v)$. Next, the image is transformed from SH domain back to the intensity signal domain $[\hat{S}(v)]$ using the harmonized SH coefficients: $\hat{S}(v) = \sum_i \sum_j \hat{C}_{ij}(v)Y_{ij}$.

Rotation invariant spherical harmonics has many advantages, the most important one being that it harmonizes the raw dMRI signal in a model-independent manner. The mapping captures only site-related differences, preserving the between-subject biological variation and fiber orientation (Karayumak et al., 2019). However, a limitation is that it requires dMRI data with similar acquisition parameters across sites. It also requires the same number of matched controls that are scanned in both reference and target sites to obtain the scale maps.

## Machine Learning

In the past decade, several diffusion data harmonization methods have been developed employing a machine learning approach, such as sparse dictionary learning (SDL) and deep learning (DL).

### Sparse Dictionary Learning (SDL)

Sparse dictionary learning is a representation learning method aiming at representing the input data as a linear combination of elements (the sparse dictionary), thus reducing the complexity of the harmonization problem (Mairal et al., 2010). The dictionary elements are small patches of spatial and angular image features (e.g., 3 × 3 × 3 × 5 voxels) that are learnt from the data itself. From a large set of random features, SDL extracts the common features with which full images can be reconstructed. The idea behind applying SDL for harmonization is that when a sparse dictionary can be constructed from data originating from multiple sites, the learnt imaging features will not include features of inter-site variability, as those are not common across the input data. Reconstructing dMRI data with a sparse dictionary, would then effectively harmonize the data (St-Jean et al., 2016; St-Jean et al., 2017).

An advantage of this method is that modeling a signal with such a sparse decomposition (sparse coding) is very effective in detecting salient regions that are related to the more informative

areas. However, a disadvantage is that, depending on the interest points and the type/resolution of the image, sometimes only a few regions are detected.

## Deep Learning (DL)

The DL approach, which is illustrated in **Figure 4**, consists of two steps: (1) Training: the learning stage in which the network parameters are optimized using the DWI from the same subjects acquired in two sites (target and reference) and (2) Inference: the trained network is applied to harmonize all subjects of the target site.

The current deep learning algorithms for diffusion data harmonization are mainly based on spherical harmonic features. The aim is to bring all the images in the same SH domain, by modifying the SH coefficients of the target data creating harmonized DWIs of the target site that are comparable to the DWIs from reference site. To achieve this, the network is trained to generate the harmonized image starting from the image acquired at a target site, using the image acquired in the reference site as ground truth, as illustrated in **Figure 4**. Hence, diffusion data from subjects that were acquired in both reference and target sites are used for training the network. Once it is trained, the inference can be done for other subjects from the target site, to create harmonized images.

Tax et al. (2019) presented a summary of four deep learning algorithms and one sparse dictionary learning harmonization algorithm used to evaluate two harmonization tasks in diffusion MRI: scanner-to-scanner mapping and angular- and spatial-resolution enhancement, i.e., mapping between standard and state-of-the-art acquisitions. Each of the algorithms was built with different net architectures and strategies. The deep learning algorithms that were evaluated by Tax et al. (2019) are: spherical harmonic network (SHNet), spherical harmonic residual network (SHResNet), spherical network (SphericalNet), and fully convolutional shuffling network (FCSNet). The used SH coefficients, on which the net is based, are obtained starting from the diffusion signal of the same subjects scanned in different scanners and with different acquisition schemes. Here we summarize some of these methods. A more extensive benchmark can be found in Tax et al. (2019).

### *Spherical Harmonic Network (SHNet)*

Spherical Harmonic Network is based on a classical Fully Connected Network (FCN) architecture, composed of a cascade of three fully connected layers, in which the rectified linear unity (ReLU) function is used as the activation function (Golkov et al., 2016; Koppers et al., 2017). Next, a batch normalization layer is used to stabilize. The different weights of the neural network layers are tuned by using paired images from different sites. The net is trained by matching data between the target site and the reference site to obtain the harmonized image. Once the network is trained, it can be used to harmonize unseen datasets from the target site. The main advantage of this network is that it is a simple FCN approach to tackle the harmonization problem. However, it might not be sufficiently sensitive to learn all the complex features of an accurate harmonization procedure.

### *Spherical Harmonic Residual Network (SHResNet)*

A Convolutional Neural Network (CNN) approach has been presented by Koppers et al. (2018). In this case, the network algorithm is based on the novel concept of residual structure by He et al. (2016). This approach is based on the difference between the input and the ground truth (target signal). The main building blocks of SHResNet are so-called functional units consisting of three convolutional layers, where each functional unit predicts the coefficients of a single SH order (Koppers et al., 2017). The main advantage of using a residual network structure consists in the robustness against the degradation problem (decrease of accuracy due to the increased network depth) and hence enabling the use of a deeper network (more convolutional layers). Nonetheless, the harmonization is done per harmonic order of the SH signal, thus, the signal from both target and reference should have the same SH orders.

### *Spherical Network (SphericalNet)*

SphericalNet is a novel deep learning approach based on spherical surface convolutions (Koppers and Merhof, 2018). It transforms the signal from SH space into spherical surface space, and performs three spherical surface convolutions. After each of these convolutions, a sigmoid activation function is applied in order to limit the signal's range between 0 and 1 (Tax et al., 2019). The signal is converted back to SH space, followed by three 3-D convolutional layers with parametric ReLU as activation. Spatial information is combined in the last convolutional layer to project neighborhood info into one voxel. The advantage of this algorithm is that it uses spherical information during spatial convolution to improve accuracy in the harmonization procedure. However, for this algorithm the intensity signal has to be transformed twice (for SH domain and then to spherical surface domain), which could introduce additional complexity to the harmonization problem.

### *Fully Convolutional Shuffling Network (FCSNet)*

Fully convolutional network is a patch-based deep learning harmonization algorithm inspired by Tanno et al. (2017). The architecture of this network contains four hidden convolutional layers with ReLU activation. Large patches are used as input, overlayed to cover the entire brain, and smaller patches are obtained as output. The last layer contains a "shuffle" operation and is composed of "skip" connections to increase the prediction accuracy. The cost function for this algorithm has two parts: channel-wise loss and loss on the function-value. The algorithm uses the patched-based fully convolutional network for diffusion data harmonization and resolution enhancement. One advantage of this approach is the use of large patches that inform about the local neighborhood and are beneficial for the harmonization procedure. On the other hand, neighborhood data could be biased and end up corrupting the harmonization algorithm.

Deep learning algorithms demonstrated the robust capability of solving non-linear problems such as data harmonization. However, some limitations are: (1) overfitting, i.e., when the model is more accurate in fitting known data but less accurate in predicting unseen data, (2) the need for a large amount of matched subjects scanned at different sites with similar

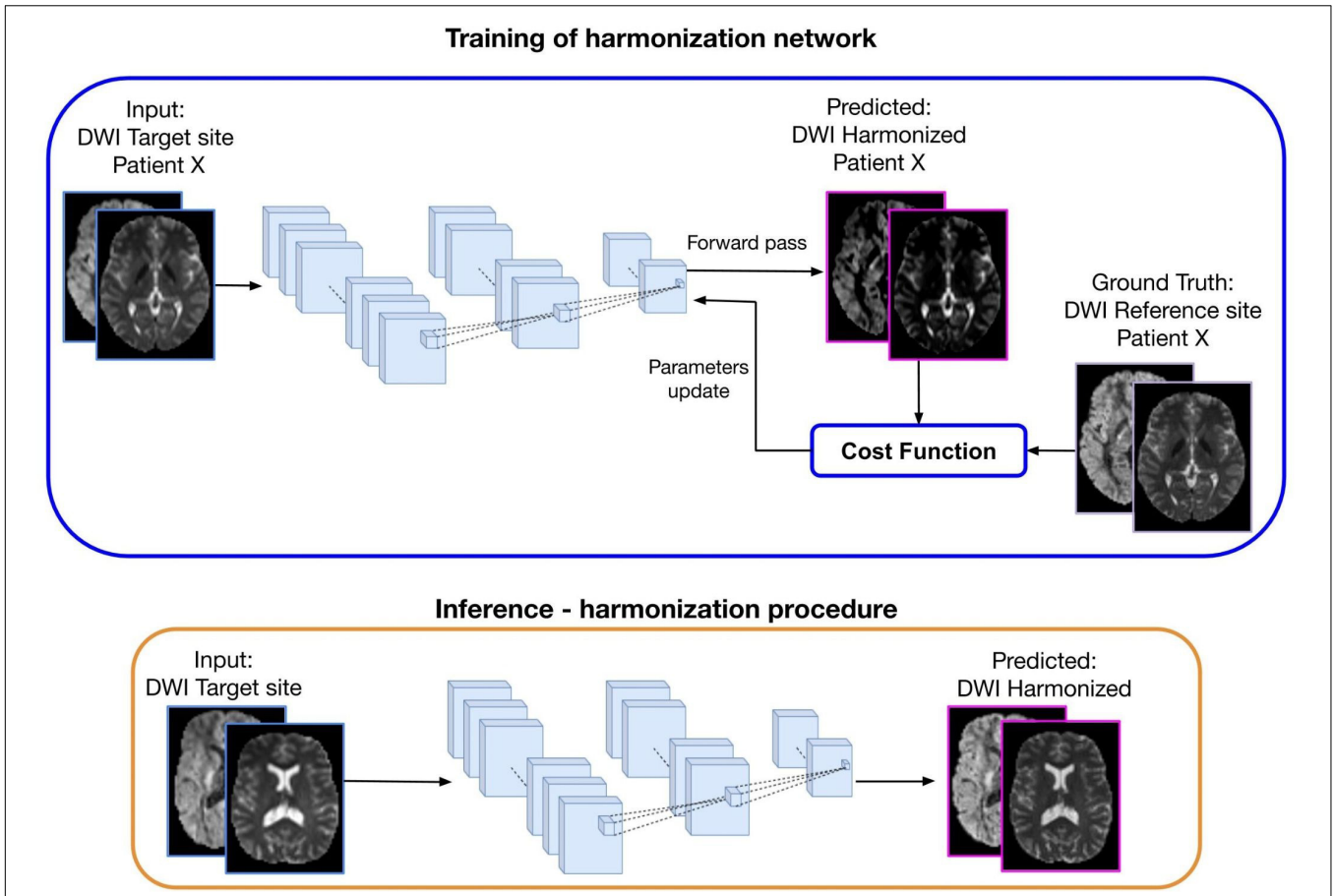


**FIGURE 4 |** Representation of a deep learning approach for diffusion data harmonization. The purpose of the method is to modify the DWI acquired at the target site, to correspond to the DWI of the same subject acquired at the reference site. In the training part, DWI from the target site is used as input and DWI from the reference site as ground truth, for patient *X*. Matched subjects are used to tune the weights of the harmonization network. During the forward phase, the network produces the predicted harmonized DWI that is compared with the corresponding expected DWI from the reference site. The difference between the predicted and the ground-truth (cost function) is back propagated into the network to update the weights in such a way that the loss decreases and the predicted harmonized DWI is closer to the ground truth. In the inference step, the trained network is used to generate the predicted harmonized DWI from unseen DWI data of the target site, which then become comparable to the DWI from the reference site.

acquisition sequences per site for training and (3) possible distortion of pathological information, if the net is trained with healthy subjects and then applied to patients.

## Method of Moments (MoM)

Method of Moments is a statistical harmonization approach that uses spherical moments to map DWI images from target to reference sites (Huynh et al., 2019). The first moment ($M_1$) corresponds to the spherical mean and the second central moment ($C_2$) corresponds to the spherical variance. The core idea is to match the spherical mean and spherical variance in order to correct for unwanted variability. Each voxel-wise n-th spherical moment ($M_\mathrm{n}$) is defined as the diffusion signal at constant $b$-value ($S_\mathrm{b}$) raised to the power of $n$ integrated over all directions $g$: $M_\mathrm{n}[S_\mathrm{b}] = \int S_\mathrm{b}^\mathrm{n}(g)dg$. MoM matches $M_1$ and $C_2$ per b-shell $b$ using the mapping function ($f_\theta$): $M_1[R_\mathrm{b}] = M_1[f_\theta(T_\mathrm{b})]$ and $C_2[R_\mathrm{b}] = C_2[f_\theta(T_\mathrm{b})]$, where $R_\mathrm{b}$ is the diffusion signal acquired at the reference site, and $T_\mathrm{b}$ the signal at the target site. Considering the mapping function as $f_{\theta=\{\alpha,\beta\}}(S) = \alpha S + \beta$, $\alpha$ and $\beta$ are the mapping coefficients calculated as $\alpha_b = \sqrt{\frac{C_2[R_\mathrm{b}]}{C_2[T_\mathrm{b}]}}$ and $\beta_\mathrm{b} = M_1[R_\mathrm{b}] - \alpha_\mathrm{b} M_1[T_\mathrm{b}]$. The MoM parameters are calculated in template space and then warped back to native space of the target subjects and applied to the DWI images. The MoM-harmonized DWI signal is $S_\mathrm{b}^\mathrm{MoM} = \alpha_\mathrm{b} S_\mathrm{b} + \beta_\mathrm{b}$.

The MoM approach is illustrated in **Figure 5**. In this method, $M_1$ and $C_2$ are computed in native space from the DWIs acquired in the reference and target sites. Next, the moment images are warped into a common space that is defined by the target data at the population level. Population moment median images across subjects are calculated for each of the moments for each of the sites. The mapping parameters ($\alpha$ and $\beta$) for the target site are obtained by matching the population median moments using the linear mapping function $f_\theta$. These parameters are warped to native space for each of the subjects of the target site and the mapping function is applied voxel-wise. Lastly, the harmonized DWI of the target data is obtained.

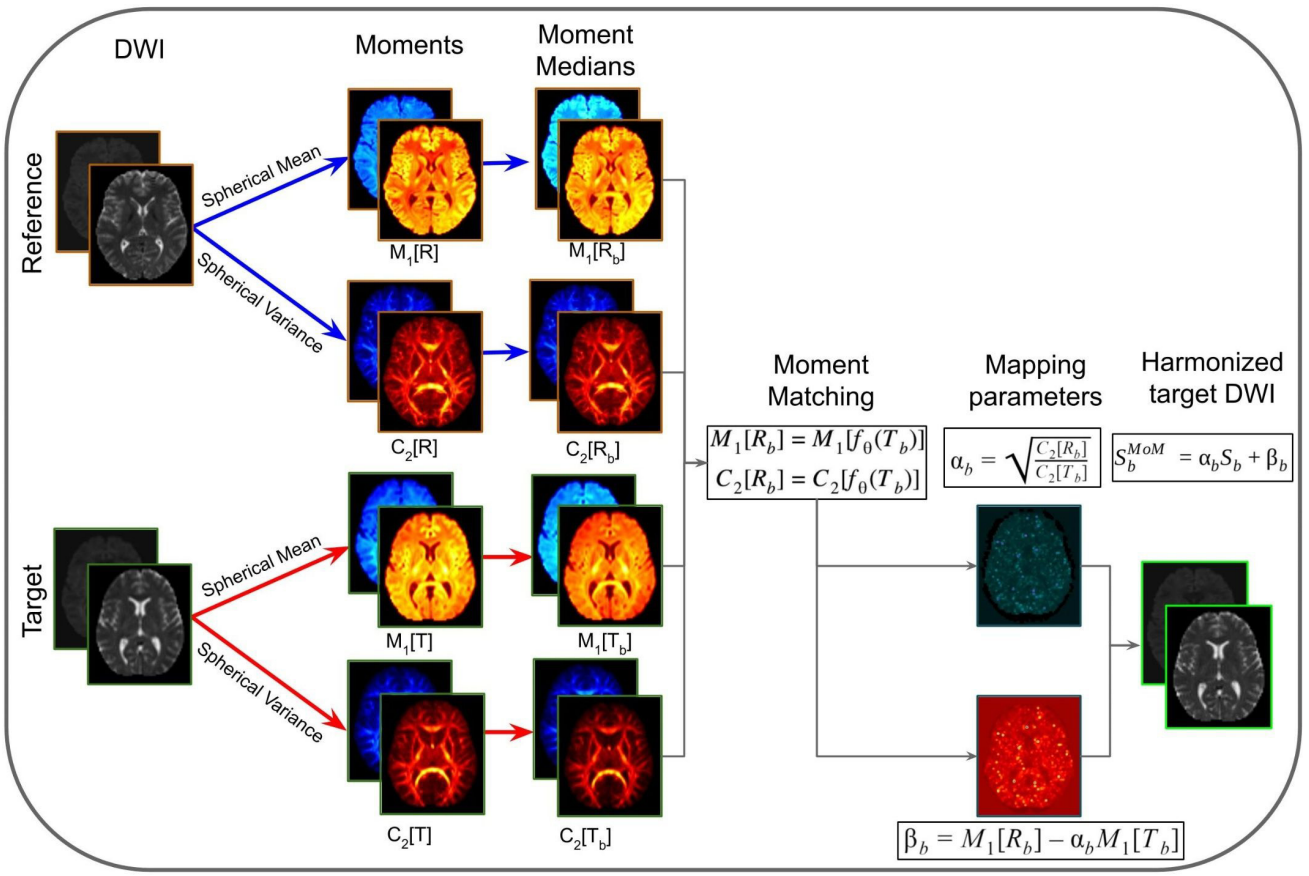


**FIGURE 5 |** Representation of the method of moments harmonization pipeline. The purpose of the method is to modify the DWI of the target site, to correspond to the DWI acquired in the reference site. Initially, the diffusion signal in the reference (*R*) and target (*T*) are used to compute spherical means ($M_1[R]$ and $M_1[T]$) and spherical variances ($C_2[R]$ and $C_2[T]$) in native space for each b-shells (*b*). The spherical moments are warped to a common space, based on the target population. Then the moment medians are calculated across subjects ($M_1[R_b]$, $C_2[R_b]$, $M_1[T_b]$, and $C_2[T_b]$). Afterward, the mapping parameters ($\alpha_b$ and $\beta_b$) are calculated per b-shell, by matching the population moments. The mapping parameters are warped to native space and applied voxel-wise to the DWI images of target site subjects, obtaining the harmonized DWI.

Advantages of the MoM are that it (1) allows direct harmonization of DWI images, without the need to represent them in any other space domain (e.g., SH space); (2) preserves directional information of the signal; (3) does not require that reference and target data have the same number of gradient directions; (4) does not require training data or matched populations with controls/patients, and (5) allows the harmonization of either a subject or a population of subjects. However, MoM as described in Huynh et al. (2019) does not harmonize multi-site data with different spatial resolution or different *b*-values. Possible solutions to cope with different spatial resolutions and different *b*-values would be to resample the reference data to the resolution of the target data, and rescale the signal, respectively, both prior to harmonization.

## DISCUSSION

Multi-center and/or longitudinal studies using diffusion MRI data are significantly affected by inter- and intra-site variability. Sources of variability include, but are not limited to, hardware, acquisition settings, reconstruction algorithms, incompatible data formats and data quality. To cope with this variability, regulations and strategies are needed to facilitate harmonization of multi-center diffusion MRI data. In that respect, MR scanner vendors and researchers have a responsibility regarding the access and storage of DWI data, and transparency on reconstruction algorithms, acquisition protocols and applied pre- and post-processing steps. Ideally, worldwide governments should ally to enforce regulations regarding calibration procedures to MR scanner vendors. The use of the same quantitative calibration phantom and a standard procedure would decrease inter-scanner variability (Keenan et al., 2017; Prohl et al., 2019).

The need for harmonization has increased with the availability of large diffusion MRI multi-center datasets. Examples of these are the Human Connectome Project (HCP[2]), the Alzheimer's

[2]https://www.humanconnectome.org/

Disease Neuroimaging Initiative (ADNI[3]), CENTER-TBI[4], and the Cross-scanner and cross-protocol diffusion MRI data harmonization (Tax et al., 2019). For performing joint analysis of data that have been acquired with multiple acquisition settings, several statistical and mathematical harmonization approaches have been developed to reduce unwanted site variability while preserving the biological variability.

To overcome the challenges with respect to joint analysis of multi-center diffusion data, the scientific community has gathered to participate in challenges on data harmonization. The Diffusion MRI Data Harmonization[5] 2017 and the Multi-shell Diffusion MRI Harmonization Challenge 2018 (MUSHAC[6]) were proposed with the aim to evaluate the performance of algorithms that enable the harmonization of DWI data. From the last challenge, Ning et al. (2019) presented a summary of results comparing the effects of DWIH methods on diffusion parametric maps. Different DWIH methods were used to harmonize the multi-shell DWI data. The algorithms range over three approaches: interpolation-based, regression-based and CNN algorithms. Diffusion parametric maps were calculated before and after the harmonization procedure, such as FA, MD, and MK. The results demonstrated that the harmonization algorithms are significantly effective in reducing the variability and maintaining the biological information.

In this paper, we have reviewed a variety of harmonization methods proposed in the literature. The decision as to which method to use depends on several aspects, such as the study design, the research question and the available data. In **Table 1**, we have categorized the reviewed methods in terms of their intrinsic properties. This categorization may help to select a harmonization method, given a certain diffusion MRI dataset and a specific research question. Additionally, **Figure 6** shows a flowchart that could provide guidance for selecting the most appropriate harmonization strategy.

For example, the flowchart can be applied to the study of Zavaliangos-Petropulu et al. (2019), who assessed the relation between diffusion MRI indices and cognitive impairment in brain aging using the ADNI3 dataset. In this study, new harmonized metrics maps (FA, MD, AD, and RD) were created using the ComBat method to remove any site-effects from the results. Following the flowchart presented in **Figure 6**, first, the research was related to the harmonization of diffusion metric maps, thus, the right segment of the chart is suggested to be followed. Next, the researchers aimed to create new harmonized maps, in this case the choice of a regression of covariates method was logical and appropriate. Along these lines, the suggested harmonization approach by our flowchart is in agreement with the decision from the authors.

In general, it is an ongoing challenge to define a gold standard for dMRI harmonization. A possible explanation for this might come from the complexity of removing the unwanted variability. The sources of unwanted variability may stem from differences in number of subjects acquired per site, MRI hardware, acquisition protocol (voxel size, repetition time, echo time, number of diffusion directions, number of b-shells, etc.), pre-processing steps and co-registration effects. In these circumstances, the preservation of expected biological variability is a useful criterion for evaluating the efficacy of harmonization methods, but this is only possible when the same subjects are scanned at different sites. When traveling human phantoms are included in the study design this provides a ground truth and allows for carefully evaluating the newly computed features and their accuracy and precision (Tax et al., 2019). However, traveling human phantoms datasets are mostly absent from a scenario of multi-center studies, where distinct subjects are scanned at different sites. Additionally, a note of caution in both cases is due here since anatomical differences or co-registration deformations (to a common space) may cause significant errors in the harmonization.

Although DPMH approaches have demonstrated their ability to harmonize diffusion metric measures for joint analysis in multi-center studies, there are some drawbacks, which can be avoided by using DWIH methods. First, DPMH methods require different transforms to harmonize each of the diffusion metrics of interest. This may have implications for multivariate analyses, as it is not guaranteed that subject-specific patterns (e.g., high FA in combination with low MD) are preserved after both metrics are harmonized separately. Second, DWIH methods do not rely on a specific diffusion model, hence unwanted variation is not propagated (and as a result made more complex) through model fitting. Moreover, any diffusion metric estimated from DWIH harmonized DWIs will automatically be harmonized as well. In this regard, DWIH approaches are more promising for reliable harmonization.

In a recent study by Cetin-Karayumak et al. (2019), DWIH was applied to harmonize diffusion MRI multi-site data prior to detection of white matter abnormalities in schizophrenia patients. RISH was retrospectively applied to DWIs of 13 different sites to remove the site-related differences. For this, a reference site was chosen and the DWI data from the other 12 sites were harmonized accordingly. The harmonization performance was evaluated in a group of matched controls, using their FA maps before and after harmonization. It was shown that the statistical differences between sites were removed and the inter-subject biological differences were preserved.

Nonetheless, many challenges remain for diffusion data harmonization in multi-center studies. Ideally, novel harmonization methods should not require training data of subjects scanned in multiple centers, and be applicable to data acquired with different spatial resolution, number of b shells, or number of diffusion gradient directions. Moreover, the availability of easily implementable methods and open-source platforms are important assets to encourage researchers to perform diffusion data harmonization in multi-center and longitudinal studies.

Furthermore, harmonization methods should be generalizable to clinical cases. Up to now, serious challenges that limit voxel-wise harmonization of DWI data of clinical patients are the co-registration requirement, since disease-related anatomical alterations may severely complicate co-registration,

[3]http://adni.loni.usc.edu/

[4]https://www.center-tbi.eu/

[5]https://projects.iq.harvard.edu/cdmri2017/challenge

[6]https://projects.iq.harvard.edu/cdmri2018/challenge

**TABLE 1 |** Overview of the harmonization methods presented in this review.

| Category | Method | | References | Statistical harmonization | Creates new harmonized images | Individual measures required | Same subjects acquired in multiple centers required | Training data required | Similar acquisition protocols required | Inter subject co-registration of DWI required | Mapping parameters on template space |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **DPMH** | | Meta-analysis | Salimi-Khorshidi et al., 2009; Teipel et al., 2012; Jahanshad et al., 2013; Kochunov et al., 2014; Zhu et al., 2019 | X | | | | | | | |
| | | Mega-analysis | | X | | X | | | | | |
| | Regression of covariates | Human-phantom based harmonization (HuP) | Pohl et al., 2016 | X | X | X | | | X | | |
| | | Hardware-phantom based harmonization (HaP) | Timmermans et al., 2019 | X | | X | | | X | | |
| | | Global Scaling (GS) | Fortin et al., 2017 | X | X | X | | | | X | |
| | | Removal of Artificial Voxel Effect by Linear Regression (RAVEL) | Fortin et al., 2016 | X | X | X | | | | X | X |
| | | Surrogate Variable Analysis (SVA) | Leek et al., 2012; Fortin et al., 2017 | X | X | X | | | | X | X |
| | | Combined association test (ComBat) | Fortin et al., 2017 | X | X | X | | | | X | X |
| **DWIH** | Rotation Invariant Spherical Harmonics (RISH) | | Mirzaalian et al., 2015, 2016, 2018; Karayumak et al., 2019 | | X | X | | X | X | X | X |
| | Machine learning | Sparse Dictionary Learning (SDL) | St-Jean et al., 2016, 2017; Tax et al., 2019 | | X | X | X | | | X | |
| | | Deep Learning (DL) | Golkov et al., 2016; Koppers et al., 2017, 2018; Tanno et al., 2017; Koppers and Merhof, 2018; Tax et al., 2019 | | X | X | X | | | X | |
| | Method of Moments (MoM) | | Huynh et al., 2019 | X | X | X | | | | X | X |

*Comparison between the methods related to: implementation of statistical harmonization, creation of new harmonized images, requirement of individual measures, requirement of images from the same subjects acquired in multiple centers, requirement of training data (i.e., matched subjects across sites are needed for obtaining the mapping between sites), requirement of similar acquisition protocols (i.e., diffusion directions, b-values, spatial resolution, TR, TE, SNR, etc.), requirement of inter subject co-registration, and implementation of mapping between sites through mapping parameters on template space. The X denotes if the method requires or performs the specific condition described in the column.*

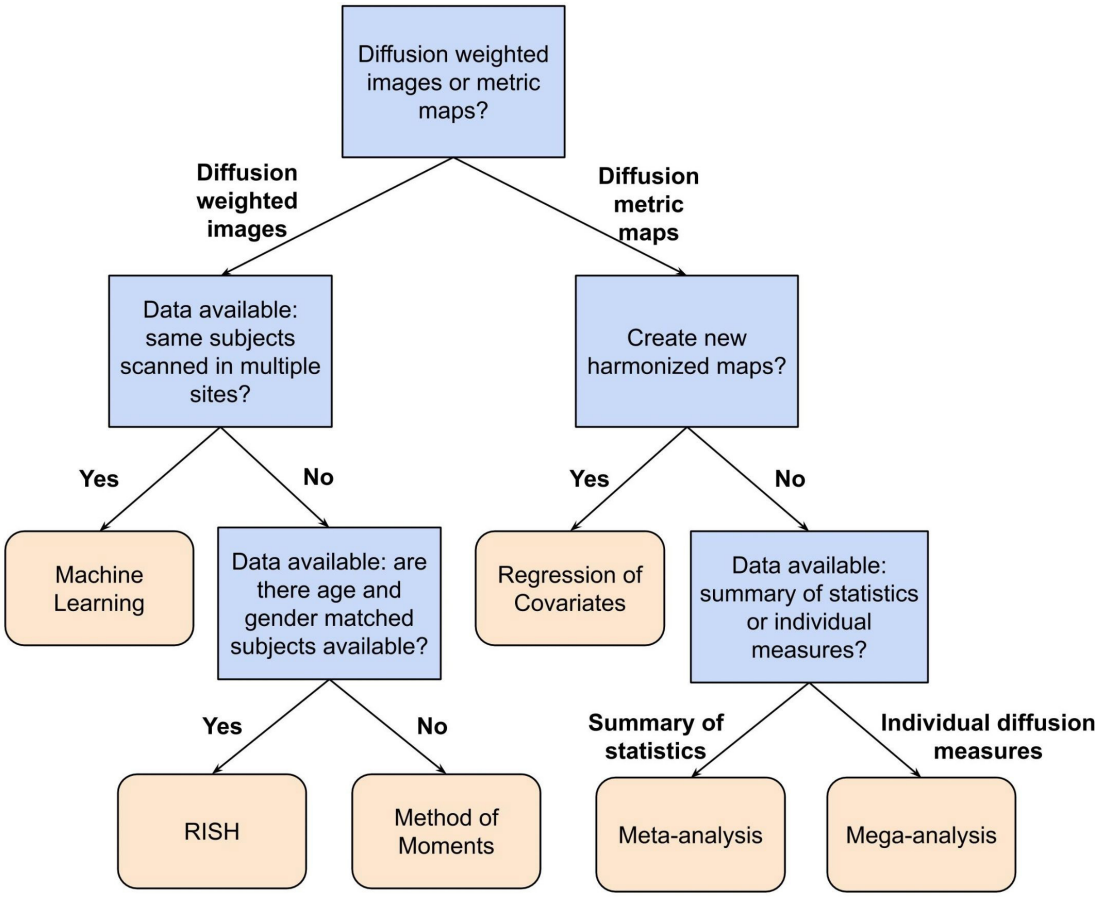


**FIGURE 6 |** Flowchart describing a possible way to select a suitable harmonization method depending on the available data and research question at hand. In this flowchart, the first question to be answered is: Do you want to harmonize the DWI or the diffusion metric maps? For harmonization of the diffusion metric maps (right segment of the flowchart), the following question is: Do you want to create new harmonized metric maps? If so, the suggested harmonization approach would be one of the regression of covariates methods. In case of a negative answer, the next question is: Do you have individual measures available or a summary of statistics? If the user has a summary of statistics, the suggestion is to use a meta-analysis approach, otherwise, if one has individual diffusion measures, the suggestion is to harmonize the data using a mega-analysis approach. On the other hand, for harmonization of DWIs (left segment of the flowchart), the next question is: Do you have DWIs of the same subjects acquired in multiple sites? In case of an affirmative answer, the suggested approach is machine learning, which comprehends deep learning and sparse dictionary learning methods. In case of a negative answer, the following question is: Do you have DWIs of a cohort of subjects that is age- and gender-matched between the sites? If the user has matched data, the RISH method is suggested. Otherwise, the method of moments is the suggested approach.

and the condition that the pathological content (e.g., diffusion properties of lesions) should be harmonized while the expected biological variability should not be affected. To overcome these limitations, the use of clinical data during the training of DWIH harmonization approaches would be valuable.

## CONCLUSION

While dMRI is routinely used in clinical workflows, comparing the signal intensity of dMRI scans across sites and over time is challenging. Harmonization methods aim to overcome this by recalibrating/recalculating either the DWI signal intensities or the resulting diffusion metrics. In this article an overview of harmonization methods in the literature was presented, covering meta- and mega-analysis, regression of covariates, rotation invariant spherical harmonics, machine learning algorithms and the method of moments. The proposed feature table and flowchart present the main characteristics of the methods, assisting in the decision of which method to use depending on the study design and the available data. Future developments of diffusion harmonization methods may benefit from focusing on DWIH approaches, avoiding unwanted variation propagates through diffusion model fitting.

## AUTHOR CONTRIBUTIONS

MP and RP wrote the manuscript with comments from TB, PVD, P-JG, BJ, AR, AdD, and JS. MP and RP contributed equally to this manuscript. All authors read and approved the final manuscript.

## FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764513. PVD and BJ are supported by the Research Foundation (FWO) Flanders, Belgium. This work was also supported by the European Space Agency (ESA) and BELSPO Prodex and the Flemish Government under the Research program Artificial Intelligence (AI) Flanders. The diffusion data used for **Figures 3**, **4**, and **5** were acquired at the United Kingdom National Facility for *In Vivo* MR Imaging of Human Tissue Microstructure located in CUBRIC funded by the EPSRC (grant EP/M029778/1), and The Wolfson Foundation. Acquisition and processing of the data was supported by a Rubicon grant from the NWO (680-50-1527), a Wellcome Trust Investigator Award (096646/Z/11/Z), and a Wellcome Trust Strategic Award (104943/Z/14/Z). This database was initiated by the 2017 and 2018 MICCAI Computational Diffusion MRI committees (Chantal Tax, Francesco Grussu, Enrico Kaden, Lipeng Ning, Jelle Veraart, Elisenda Bonet-Carne, and Farshid Sepehrband) and CUBRIC, Cardiff University (Chantal Tax, Derek Jones, Umesh Rudrapatna, John Evans, Greg Parker, Slawomir Kusmia, Cyril Charron, and David Linden).

## REFERENCES

Alexander, A. L., Lee, J. E., Wu, Y. C., and Field, A. S. (2006). Comparison of diffusion tensor imaging measurements at 3.0 *T versus* 1.5 T with and without parallel imaging. *Neuroimag. Clin. N. Am.* 16, 299–309. doi: 10.1016/j.nic.2006.02.006

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044.

Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage* 20, 1052–1063. doi: 10.1016/S1053-8119(03)00435-X

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). "Power failure: why small sample size undermines the reliability of neuroscience": Erratum. *Nat. Rev. Neurosci.* 14:442.

Cannon, T. D., Sun, F., McEwen, S. J., Papademetris, X., He, G., van Erp, T. G. M., et al. (2014). Reliability of neuroanatomical measurements in a multisite longitudinal study of youth at risk for psychosis. *Hum. Brain Mapp.* 35, 2424–2434. doi: 10.1002/hbm.22338

Cetin-Karayumak, S., Di Biase, M. A., Chunga, N., Reid, B., Somes, N., Lyall, A. E., et al. (2019). White matter abnormalities across the lifespan of schizophrenia: a harmonized multi-site diffusion MRI study. *Mol. Psychiatry* doi: 10.1038/s41380-019-0509-y [Epub ahead of print]

Forsyth, J. K., McEwen, S. C., Gee, D. G., Bearden, C. E., Addington, J., Goodyear, B., et al. (2014). Reliability of functional magnetic resonance imaging activation during working memory in a multi-site study: analysis from the North American prodrome longitudinal study. *Neuroimage* 97, 41–52. doi: 10.1016/j.neuroimage.2014.04.027

Fortin, J. P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., et al. (2017). Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161, 149–170. doi: 10.1016/j.neuroimage.2017.08.047

Fortin, J. P., Sweeney, E. M., Muschelli, J., Crainiceanu, C. M., and Shinohara, R. T. (2016). Removing inter-subject technical variability in magnetic resonance imaging studies. *Neuroimage* 132, 198–212. doi: 10.1016/j.neuroimage.2016.02.036

Fox, R. J., Sakaie, K., Lee, J.-C., Debbins, J. P., Liu, Y., Arnold, D. L., et al. (2012). A validation study of multicenter diffusion tensor imaging: reliability of fractional anisotropy and diffusivity values. *Am. J. Neuroradiol.* 33, 695–700. doi: 10.3174/ajnr.A2844

Golkov, V., Dosovitskiy, A., Sperl, J. I., Menzel, M. I., Czisch, M., Samann, P., et al. (2016). Q-space deep learning: twelve-fold shorter and model-free diffusion MRI scans. *IEEE Trans. Med. Imaging* 35, 1344–1351. doi: 10.1109/tmi.2016.2551324

Grech-Sollars, M., Hales, P. W., Miyazaki, K., Raschke, F., Rodriguez, D., Wilson, M., et al. (2015). Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. *NMR Biomed.* 28, 468–485. doi: 10.1002/nbm.3269

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, 770–778. doi: 10.1109/CVPR.2016.90

Huynh, K. M., Chen, G., Wu, Y., Shen, D., and Yap, P. (2019). Multi-site harmonization of diffusion MRI data via method of moments. *IEEE Trans. Med. Imaging* 38, 1599–1609. doi: 10.1109/tmi.2019.2895020

Jahanshad, N., Kochunov, P. V., Sprooten, E., Mandl, R. C., Nichols, T. E., Almasy, L., et al. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA–DTI working group. *Neuroimage* 81, 455–469. doi: 10.1016/j.neuroimage.2013.04.061

Jenkins, J., Chang, L. C., Hutchinson, E., Irfanoglu, M. O., and Pierpaoli, C. (2016). "Harmonization of methods to facilitate reproducibility in medical data processing: applications to diffusion tensor magnetic resonance imaging," in *Proceedings–2016 IEEE International Conference on Big Data, Big Data 2016*, Washington, DC, 3992–3994. doi: 10.1109/BigData.2016.7841086

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/J.NEUROIMAGE.2011.09.015

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037

Jones, D. K. (2010). Precision and accuracy in diffusion tensor magnetic resonance imaging. *Topics Magn. Reson. Imaging* 21, 87–99. doi: 10.1097/RMR.0b013e31821e56ac

Jovicich, J., Barkhof, F., Babiloni, C., Herholz, K., Mulert, C., van Berckel, B. N. M., et al. (2019). Harmonization of neuroimaging biomarkers for neurodegenerative diseases: a survey in the imaging community of perceived barriers and suggested actions. *Alzheimer's Dementia* 11, 69–73. doi: 10.1016/j.dadm.2018.11.005

Jovicich, J., Marizzoni, M., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., et al. (2014). Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. *Neuroimage* 101, 390–403. doi: 10.1016/j.neuroimage.2014.06.075

Karayumak, S. C., Bouix, S., Ning, L., James, A., Crow, T., Shenton, M., et al. (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

Keenan, K. E., Ainslie, M., Barker, A. J., Boss, M. A., Cecil, K. M., Charles, C., et al. (2017). Quantitative magnetic resonance imaging phantoms: a review and the need for a system phantom. *Magn. Reson. Med.* 79, 48–61. doi: 10.1002/mrm.26982

Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., et al. (2018). Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI working group. *Mol. Psychiatry* 23, 1261–1269. doi: 10.1038/mp.2017.170

Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. (2010). elastix: a toolbox for intensity- based medical image registration. *IEEE Trans. Med. Imaging* 29, 196–205. doi: 10.1109/TMI.2009.2035616

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T. E., Mandl, R. C., Almasy, L., et al. (2014). Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and megaanalytical approaches for data pooling. *Neuroimage* 95, 136–150. doi: 10.1016/j.neuroimage.2014.03.033

Koppers, S., Bloy, L., Berman, J. I., Tax, C. M. W., Edgar, J. C., and Merhof, D. (2018). "Spherical harmonic residual network for diffusion signal harmonization," in *Computational Diffusion MRI*, eds E. Bonet-Carne, F. Grussu, L. Ning, F. Sepehrband, C. M. W. Tax (Berlin: Springer).

Koppers, S., Haarburger, C., and Merhof, D. (2017). "Diffusion MRI signal augmentation: from single shell to multi shell with deep learning," in *Proceedings of the Computational Diffusion MRI: MICCAI Workshop*, Athens, 61–70. doi: 10.1007/978-3-319-54130-3_5

Koppers, S., and Merhof, D. (2018). *DELIMIT PyTorch - An extension for Deep Learning in Diffusion Imaging*. Available online at: http://arxiv.org/abs/1808.01517 (accessed June 20, 2019).

Kumar, R., Gupta, R. K., Husain, M., Chaudhry, C., Srivastava, A., Saksena, S., et al. (2009). Comparative evaluation of corpus callosum DTI metrics in acute mild and moderate traumatic brain injury: its correlation with neuropsychometric test. *Brain Injury* 23, 675–685. doi: 10.1080/02699050903014915

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. doi: 10.1093/bioinformatics/bts034

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 11, 733–739. doi: 10.1038/nrg2825

Mahoney, C. J., Simpson, I. J. A., Nicholas, J. M., Fletcher, P. D., Downey, L. E., Golden, H. L., et al. (2015). Longitudinal diffusion tensor imaging in frontotemporal dementia. *Ann. Neurol.* 77, 33–46. doi: 10.1002/ana.24296

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* 11, 19–60. doi: 10.1145/1756006.1756008

Mirzaalian, H., de Pierrefeu, A., Savadjiev, P., Pasternak, O., Bouix, S., Kubicki, M., Rathi, Y. (2015). Harmonizing Diffusion MRI Data Across Multiple Sites and Scanners. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, eds N. Navab, J. Hornegger, W. M. Wells, & A. Frangi (Cham: Springer International Publishing), doi: 10.1007/978-3-319-24553-9_2

Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., et al. (2016). Inter-site and inter-scanner diffusion MRI data harmonization. *Neuroimage* 135, 311–323. doi: 10.1016/j.neuroimage.2016.04.041

Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., et al. (2018). Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging Behav.* 12, 284–295. doi: 10.1007/s11682-016-9670-y

Nencka, A. S., Meier, T. B., Wang, Y., Muftuler, L. T., Wu, Y.-C., Saykin, A. J., et al. (2017). Stability of MRI metrics in the advanced research core of the NCAA-DoD concussion assessment, research and education (CARE) consortium. *Brain Imaging Behav.* 12, 1121–1140. doi: 10.1007/s11682-017-9775-y

Ni, H., Kavcic, V., Zhu, T., Ekholm, S., and Zhong, J. (2006). Effects of number of diffusion gradient directions on derived diffusion tensor imaging indices in human brain. *Am. J. Neuroradiol.* 27, 1776–1781.

Ning, L., Bonet-Carne, E., Grussu, F., Sepehrband, F., Kaden, E., Veraart, J., et al. (2019). "Muti-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC): progress and results," in *Proceedings of the Computational Diffusion MRI: International MICCAI Workshop*, eds L. Ning, C. M. W. Tax, F. Grussu, E. Bonet-Carne, & F. Sepehrband (Cham: Springer), 217–224. doi: 10.1007/978-3-030-05831-9_18

Nyholm, T., Jonsson, J., Söderström, K., Bergström, P., Carlberg, A., Frykholm, G., et al. (2013). Variability in prostate and seminal vesicle delineations defined on magnetic resonance images, a multi-observer, -center and -sequence study. *Radiat. Oncol.* 8:126. doi: 10.1186/1748-717X-8-126

Palacios, E. M., Martin, A. J., Boss, M. A., Ezekiel, F., Chang, Y. S., Yuh, E. L., et al. (2017). Toward precision and reproducibility of diffusion tensor imaging: a multicenter diffusion phantom and traveling volunteer study. *Am. J. Neuroradiol.* 38, 537–545. doi: 10.3174/ajnr.A5025

Petitti, D. B. (1994). *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*. New York, NY: Oxford University Press.

Pfefferbaum, A., Adalsteinsson, E., and Sullivan, E. V. (2003). Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J. Magn. Reson. Imaging* 18, 427–433. doi: 10.1002/jmri.10377

Pohl, K. M., Sullivan, E. V., Rohlfing, T., Chu, W., Kwon, D., Nichols, B. N., et al. (2016). Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. *Neuroimage* 130, 194–213. doi: 10.1016/j.neuroimage.2016.01.061

Pourhoseingholi, M. A., Baghestani, A. R., and Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterol. Hepatol.* 5, 79–83. doi: 10.22037/ghfbb.v5i2.246

Prohl, A. K., Scherrer, B., Tomas-Fernandez, X., Filip-Dhima, R., Kapur, K., Velasco-Annis, C., et al. (2019). Reproducibility of structural and diffusion tensor imaging in the TACERN multi-center study. *Front. Integr. Neurosci.* 13:24. doi: 10.3389/fnint.2019.00024

Pullens, P., Bladt, P., Sijbers, J., Maas, A. I. R., and Parizel, P. M. (2017). Technical Note: a safe, cheap, and easy-to-use isotropic diffusion MRI phantom for clinical and multicenter studies. *Med. Phys.* 44, 1063–1070. doi: 10.1002/mp.12101

Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., and Nichols, T. E. (2009). Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage* 45, 810–823. doi: 10.1016/j.neuroimage.2008.12.039

Shamonin, D. P., Bron, E. E., Lelieveldt, B. P., Smits, M., Klein, S., Staring, M., et al. (2014). Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front. Neuroinform.* 7:50. doi: 10.3389/fninf.2013.00050

Smith, S. M., and Nichols, T. E. (2018). Statistical Challenges in "Big Data". *Hum. Neuroimag. Neuron* 97, 263–268. doi: 10.1016/j.neuron.2017.12.018

St-Jean, S., Coupé, P., and Descoteaux, M. (2016). Non local spatial and angular matching: enabling higher spatial resolution diffusion MRI datasets through adaptive denoising. *Med. Image Anal.* 32, 115–130. doi: 10.1016/j.media.2016.02.010

St-Jean, S., Viergever, M., and Leemans, A. (2017). "A unified framework for upsampling and denoising of diffusion MRI data," in *Proceedings of the 25th Annual Meeting of ISMRM*, Honolulu, HI, 3533.

Tanno, R., Worrall, D. E., Ghosh, A., Kaden, E., Sotiropoulos, S. N., Criminisi, A., et al. (2017). "Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super-resolution," *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science*, eds M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. Collins, S. Duchesne (Cham: Springer), Vol 10433, 611–619. doi: 10.1007/978-3-319-66182-7_70

Tax, C. M., Grussu, F., Kaden, E., Ning, L., Rudrapatna, U., John Evans, C., et al. (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

Teipel, S. J., Reuter, S., Stieltjes, B., Acosta-Cabronero, J., Ernemann, U., Fellgiebel, A., et al. (2011). Multicenter stability of diffusion tensor imaging measures: a European clinical and physical phantom study. *Psychiatry Res. Neuroimaging* 194, 363–371. doi: 10.1016/j.pscychresns.2011.05.012

Teipel, S. J., Wegrzyn, M., Meindl, T., Frisoni, G., Bokde, A. L. W., Fellgiebel, A., et al. (2012). Anatomical MRI and DTI in the diagnosis of Alzheimer's disease: a european multicenter study. *J. Alzheimer's Dis.* 31, S33–S47. doi: 10.3233/jad-2012-112118

Timmermans, C., Smeets, D., Verheyden, J., Terzopoulos, V., Anania, V., Parizel, P. M., et al. (2019). Potential of a statistical approach for the standardization of multicenter diffusion tensor data: a phantom study. *J. Magn. Reson. Imaging* 49, 955–965. doi: 10.1002/jmri.26333

Tong, Q., He, H., Gong, T., Li, C., Liang, P., Qian, T., et al. (2019). Reproducibility of multi-shell diffusion tractography on traveling subjects: a multicenter study prospective. *Magn. Reson. Imaging* 59, 1–9. doi: 10.1016/j.mri.2019.02.011

Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., et al. (2015). Predicting advertising success beyond traditional measures: new insights from neurophysiological methods and market response modeling. *J. Mark. Res.* 52, 436–452. doi: 10.1509/jmr.13.0593

Vollmar, C., O'Muircheartaigh, J., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., et al. (2010). Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *Neuroimage* 51, 1384–1394. doi: 10.1016/j.neuroimage.2010.03.046

Walker, L., Curry, M., Nayak, A., Lange, N., and Pierpaoli, C. (2013). A framework for the analysis of phantom data in multicenter diffusion tensor imaging studies. *Hum. Brain Mapp.* 34, 2439–2454. doi: 10.1002/hbm.22081

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage* 21, 1732–1747. doi: 10.1016/j.neuroimage.2003.12.023

Worsley, A. (2002). Nutrition knowledge and food consumption: can nutrition knowledge change food behaviour? *Asia Pacific J. Clin. Nutr.* 11, S579–S585. doi: 10.1046/j.1440-6047.11.supp3.7.x

Zavaliangos-Petropulu, A., Nir, T. M., Thomopoulos, S. I., Reid, R. I., Bernstein, M. A., Borowski, B., et al. (2019). Diffusion MRI indices and their relation to cognitive impairment in brain aging: the updated multi-protocol approach in ADNI3. *Front. Neuroinform.* 13:2. doi: 10.3389/fninf.2019.00002

Zhu, A. H., Moyer, D. C., Nir, T. M., Thompson, P. M., and Jahanshad, N. (2019). *Challenges and Opportunities in dMRI Data Harmonization. In Computational Diffusion MRI*. Berlin: Springer International Publishing, 157–172. doi: 10.1007/978-3-030-05831-9_13

Zhu, T., Hu, R., Qiu, X., Taylor, M., Tso, Y., Yiannoutsos, C., et al. (2011). Measurements?: a diffusion phantom and human brain study. *Neuroimage* 56, 1398–1411. doi: 10.1016/j.neuroimage.2011.02.010.Quantification

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Pinto, Paolella, Billiet, Van Dyck, Guns, Jeurissen, Ribbens, den Dekker and Sijbers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*