Harmonization of Structural Brain Connectivity through Distribution Matching

Zhen Zhou,* Bruce Fischl, and Iman Aganj*

Athinoula A. Martinos Center for Biomedical Imaging, Radiology Department,

Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

{zzhou22, bfischl, iaganj}@mgh.harvard.edu

*Corresponding authors

September 5, 2024

Keywords: Diffusion MRI, harmonization, distribution matching, structural brain connectivity, and connectome

Abstract: The increasing prevalence of multi-site diffusion-weighted magnetic resonance imaging (dMRI) studies potentially offers enhanced statistical power for investigating brain structure. However, these studies face challenges due to variations in scanner hardware and acquisition protocols. While several methods exist for dMRI data harmonization, few specifically address structural brain connectivity. We introduce a new distribution-matching approach to harmonizing structural brain connectivity across different sites and scanners. We evaluate our method using structural brain connectivity data from two distinct datasets of OASIS-3 and ADNI-2, comparing its performance to the widely used ComBat method. Our approach is meant to align the statistical properties of connectivity data from these two datasets. We examine the impact of harmonization on the correlation of brain connectivity with the Mini-Mental State Examination score and age. Our results demonstrate that our distribution-matching technique more effectively harmonizes structural brain connectivity, often producing stronger and more significant correlations compared to ComBat. Qualitative assessments illustrate the desired distributional alignment of ADNI-2 with OASIS-3, while quantitative evaluations confirm robust performance. This work contributes to the growing

field of dMRI harmonization, potentially improving the reliability and comparability of structural connectivity studies that combine data from different sources in neuroscientific and clinical research.

1 Introduction

Diffusion-weighted magnetic resonance imaging (dMRI) is a powerful non-invasive technique for probing the microstructure of biological tissue, particularly the brain white matter (Tournier et al., 2011). By measuring the diffusion of water molecules, dMRI allows us to infer the organization and integrity of neural tissue, which makes it instrumental in both neuroscientific research and clinical applications. Structural brain connectivity, typically represented as networks derived from dMRI fiber tracking (tractography) (Behrens et al., 2007; Mori et al., 1999), provides crucial insights into the architecture of white-matter pathways and the organization of communication pathways in the brain (Bazinet et al., 2023; Yeh et al., 2021).

dMRI has been widely used to study brain development, aging, as well as various neurological and psychiatric disorders (Beck et al., 2021; Frau-Pascual et al., 2021; Pines et al., 2020; Wheeler & Voineskos, 2014), with recent years seeing an increase in multi-site dMRI studies to investigate brain disorders on a larger scale, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack Jr et al., 2015; Weiner et al., 2015), Huntington's disease research (Magnotta et al., 2012), and the Parkinson's Progression Markers Initiative (Marek et al., 2011). These studies offer the potential for increased statistical power and the ability to detect subtle effects that may not be apparent in smaller, single-site studies. However, the variability in scanner hardware, acquisition protocols, and processing methods across different sites can introduce unwanted variability in the data, potentially confounding biological effects of interest. This challenge has highlighted the critical need for robust methods to harmonize dMRI data across different scanners, protocols, and populations (Zhu et al., 2019).

Several approaches have been proposed to address the harmonization problem. One widely used method is ComBat, initially designed for genomics data and later adapted for neuroimaging (Johnson et al., 2007). ComBat and its variants have been successfully applied to a variety of neuroimaging studies to harmonize diffusion tensor imaging (DTI) measures, cortical thickness, regional volumetric measures, and functional connectivity properties (Fortin et al., 2017, 2018; Pomponio et al., 2020; Yu et al., 2018; Zhou et al., 2022, 2023). However, a limitation of ComBat is that its optimization procedure assumes a normal distribution for the data, which may not necessarily be appropriate for all types of neuroimaging measures or data (Johnson et al., 2007; Pinto et al., 2020; Pomponio et al., 2020).

Another notable method designed explicitly for dMRI data harmonization is the Rotation Invariant Spherical Harmonics (RISH) approach, introduced by Mirzaalian et al. (2016). RISH features are computed from spherical harmonic decompositions of the diffusion signal, with linear mappings used to harmonize these features across sites. Karayumak et al. (2019) extended this method for varying acquisition parameters, while De Luca et al. (2022) validated it on multi-shell dMRI data from three sites. Building on RISH, several deep-learning methods have been developed to leverage Convolutional Neural Networks to harmonize dMRI data (Koppers et al., 2017, 2019; Tax et al., 2019), preserving rotational invariance while leveraging the power of deep-learning architectures to learn complex mappings between scanner-specific features. More recently, an adaptive template estimation approach combined with RISH features was proposed to better account for inter-subject anatomical variability when harmonizing voxel-wise dMRI signals or surface-based features across different scanners (Xia & Shi, 2022, 2024). The RISH method requires well-matched control subjects (16-20 per site) to learn harmonization parameters, which can be challenging for retrospective or rare-condition studies. Furthermore, harmonizing diffusion images early in the pipeline necessitates rerunning subsequent processing steps, which increases the computational time.

A few other notable approaches to dMRI harmonization are as follows. The Method of Moments (MoM), proposed by Huynh et al. (2019), matches the first and second moments (spherical mean and variance) of the diffusion signal across sites using a linear mapping function. To capture all sources of scanner-specific variations, Zhang et al. (2023) proposed RELIEF (REmoval of Latent Inter-scanner Effects through Factorization), which uses a structured multivariate approach based on linked matrix factorization to model and remove sources of scanner effects such as scanner-specific means, variances, and latent patterns. Moyer et al. (2020) proposed an unsupervised approach using variational autoencoders to learn scanner-invariant representations, which can map dMRI data between different scanners and protocols without requiring paired training data, while preserving biologically meaningful information.

While these methods have shown success in harmonizing the diffusion signal and derived measures such as fractional anisotropy and mean diffusivity, fewer approaches have been developed explicitly for harmonizing structural brain connectivity (a.k.a. the connectome) (Bazinet et al., 2023). Importantly, structural connectivity may be particularly sensitive to scanner and protocol differences (Panman et al., 2019; Zhu et al., 2019), as well as the parameters specified for parcellation and tractography (Sotiropoulos & Zalesky, 2019; Yeh et al., 2021), highlighting the need for robust harmonization methods specifically tailored to the quantified structural connectivity (Kurokawa et al., 2021; Onicas et al., 2022; Patel et al., 2024). To address the gap in harmonization methods for structural connectivity, we propose a new approach, based on distribution matching, to harmonize structural brain connectivity across different sites and scanners. Distribution matching aligns the statistical properties of different datasets to minimize biases and

improve comparability (Bishop & Nasrabadi, 2006; Evans & Rosenthal, 2004). This process aligns the data distributions using methods ranging from simple linear transformations to complex techniques such as quantile normalization or histogram matching. By reducing dataset-specific variabilities, distribution matching enhances the reliability of subsequent analyses when integrating data from various sources. In a previous study, structural MRI images were harmonized across different scanners and sites by aligning voxel intensity distributions (Wrobel et al., 2020), demonstrating improved reduction of scanner-related variability while preserving biological differences in multi-site neuroimaging studies compared to existing techniques. In this study, we validated our method by assessing its performance in harmonizing multi-site structural brain connectivity data while comparing it with the ComBat approach.

The rest of the paper is organized as follows. Section 2 introduces the two datasets, the distribution-matching approach, and related numerical algorithms. Experimental results are presented in Section 3 to qualitatively and quantitatively demonstrate the ability of the proposed method to harmonize structural brain connectivity in comparison with other methods. Finally, discussions and conclusions are provided in Section 4.

2 Materials and methods

2.1 Datasets

We used the following two public dMRI datasets in our analysis, with the number of subjects indicating those processed and included in our study (see Table 1): the third phase of the Open Access Series of Imaging Studies (OASIS-3) (LaMontagne et al., 2019) comprising 761 cognitively normal and AD subjects, and the second phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI-2) (Beckett et al., 2015) including 209 participants ranging from cognitively normal individuals to those diagnosed with AD. Compared to a previous study where we used these two datasets (Aganj et al., 2023), here we excluded 10 and 8 outlier subjects from OASIS-3 and ADNI-2, respectively, during quality check. We included the Mini-Mental State Examination (MMSE) score and age as non-MRI variables in our correlation analyses.

2.2 Data processing

Anatomical MR images from the two datasets were processed using FreeSurfer (Fischl, 2012). We included each subject only once, specifically the earliest visit containing dMRI (which was often the

Table 1: Demographic information and scanner types of the two datasets

Dataset	OASIS-3	ADNI-2
Number of subjects	761	209
Percentage of males	45.0%	55.5%
Age (mean $\pm$ std)	69.4 $\pm$ 9.0	73.5 $\pm$ 6.9
MMSE (mean $\pm$ std)	28.2 $\pm$ 2.7	27.2 $\pm$ 2.6
Scanner	Siemens (Trio Tim & Biograph mMR)	GE

baseline visit), to maintain the independence of our data points through a cross-sectional study design. We then executed the FreeSurfer dMRI processing pipeline, incorporating commands from the FMRIB Software Library (FSL) (Jenkinson et al., 2012). This process involved propagating 85 automatically segmented cortical and subcortical regions from the structural to the diffusion space using boundary-based image registration (Greve & Fischl, 2009).

Next, we used our publicly available toolbox (www.nitrc.org/projects/csaodf-hough) to reconstruct the diffusion orientation distribution function in constant solid angle (CSA-ODF) (Aganj et al., 2010) and run Hough-transform global probabilistic tractography (Aganj et al., 2011) to generate an optimal (highest-score) streamline passing through each of the 10,000 seed points for each subject. We computed a symmetric structural connectivity matrix with non-negative elements for each subject by summing the tracts passing through each pair of ROIs weighted by the tract score, which is a function of the ODFs and generalized fractional anisotropy (Aganj et al., 2011). We then augmented the matrices with indirect connections using the mathematical calculation of the electric conductance (Aganj et al., 2014), resulting in a new matrix that reflects multi-synaptic pathways. More details are provided in our previous study (Aganj et al., 2023).

We transformed the connectivity values c (each element in the connectivity matrix) into the logarithmic space as $c \leftarrow \log(1 + c)$. This transformation helps to reduce the impact of outliers and enables more robust statistical analysis.

2.3 Distribution matching

The goal of distribution matching is to align the statistical properties of different datasets, ensuring that they share similar distributions. The positive values in our structural connectivity matrices are lower-bounded by zero, peak at some positive value, and have a tail with no theoretical upper bound. This led us to choose the gamma distribution as the appropriate model to represent our connectivity data, as it is well-suited for modeling positive continuous data with a skewed distribution, which is typical for

structural connectivity values. Characterized by its shape and scale parameters, the gamma distribution is formulated as:

$$g_{\beta,\gamma}(c) = \frac{1}{\Gamma(\gamma)\beta^{\gamma}} c^{\gamma-1} e^{-\frac{c}{\beta}}. \tag{1}$$

In our case, $c \geq 0$ is the structural connectivity value, $\beta > 0$ scales c hence named the scale parameter, $\gamma \geq 1$ is the shape parameter, and Γ is the gamma function ($\Gamma(\gamma) = (\gamma - 1)!$ for integer γ). Note that the distribution domain has a lower bound of zero, but is unbounded in the positive direction. Raw (original unaugmented) structural connectivity matrices often contain numerous zero values, indicating the absence of direct connections between certain brain regions. Accordingly, we formulated the probability density function (PDF) of each element in our original (unaugmented) structural connectivity matrix as a combination of a Dirac delta function at the origin and the gamma function, as follows:

$$f(c) \coloneqq \lambda\delta(c) + (1 - \lambda)g_{\beta,\gamma}(c), \tag{2}$$

where $\delta(c) = 0$ for $c \neq 0$ and $\int_{\mathbb{R}} \delta(c)\mathrm{d}c = 1$, and $0 \leq \lambda \leq 1$ represents the portion of zeros in a connectivity value across all subjects.

The above PDF describes the likelihood of the continuous random variable (structural connectivity) taking on the specific value c. The cumulative distribution function (CDF), $F(c) \coloneqq \int_{-\infty}^{c} f(c')\mathrm{d}c'$, on the other hand, represents the probability of connectivity taking a value no larger than c. By applying the *inverse* CDF, also known as the quantile function, we will transform the data from one distribution to match another, thereby achieving distributional alignment.

Given two datasets, collected at a reference site R and a new site N, with CDFs F_R and F_N, respectively, we want to calculate the transformation $T_{N\to R}$ that takes a connectivity value from the new site, c_N, and returns a value harmonized in the reference site, $c_R = T_{N\to R}(c_N)$. The harmonization task is to find the value c_R that shares the same percentile in the reference site as that of c_N in the new site, which is accomplished by subsequently applying the CDF of the new site and then the inverse CDF of the reference site to the connectivity value, i.e., $c_R = T_{N\to R}(c_N) = F_R^{-1}(F_N(c_N))$, or:

$$T_{N\to R} = F_R^{-1} \circ F_N. \tag{3}$$

For our distribution model, with the PDF in Eq. (2), the CDF is calculated as:

$$F(c) = \lambda u(c) + (1-\lambda) G_{\beta,\gamma}(c), \tag{4}$$

where u is the Heaviside step function ($u(c)$ is 0 for $c < 0$ and 1 otherwise) and $G_{\beta,\gamma}$ is the CDF corresponding to $g_{\beta,\gamma}$. The inverse of the CDF in Eq. (4), which is used in Eq. (3), has the following closed form:

$$F^{-1}(p) = \begin{cases} 0, & if\ p < \lambda \\ G^{-1}_{\beta,\gamma}\left(\frac{p-\lambda}{1-\lambda}\right), & if\ p \geq \lambda \end{cases}. \tag{5}$$

Applying $T_{N \to R}$ to a connectivity value from the new site harmonizes it into a value that preserves its percentile in the distribution of the reference site, thereby achieving distributional alignment.

Our study uses OASIS-3 as the reference site and ADNI-2 as the new site. Analyzing the brain connections between each pair of the 85 cortical and subcortical regions results in 3570 unique connectivity values. We harmonized each original connectivity value (across all subjects) independently, and then re-augmented the harmonized values (see Section 2.2). After harmonization, we undid the $\log$ transformation as $c \leftarrow e^c - 1$ to ensure the interpretability of the results.

In addition, we performed internal harmonization within the OASIS-3 dataset that includes data from two types of Siemens MRI scanners (655 subjects scanned by Trio Tim and 106 subjects by Biograph mMR). For the OASIS-3 internal harmonization, we considered subjects scanned by Trio Tim as the reference site while the others were treated as the new site.

In both aforementioned experiments, for comparison, we also harmonized the structural brain connectivity data using three variations of ComBat: (1) including both sex and age as covariates, (2) including only sex as a covariate, and (3) without including either age or sex as a covariate. In contrast to the proposed distribution matching, ComBat does not guarantee the non-negativity of the connectivity values, which is required for augmentation. Consequently, to harmonize augmented connectivity with ComBat, we applied ComBat directly to augmented connectivity (rather than applying it to original connectivity and re-augmenting the results, as we did with distribution matching). We systematically compared the results of ComBat variations with those obtained from our distribution matching method.

2.4 Correlation analysis

We used MMSE and age for almost all subjects to investigate how they correlated with structural brain connectivity before and after harmonization and determine if harmonization improved the outcomes of correlation analysis. For robustness, we computed the correlation on a transformed version of the connectivity value (each element in the original or augmented connectivity matrix) as $c \leftarrow 1 - e^{-\frac{c}{\bar{c}}}$, where $\bar{c}$ is the cross-subject average of c, thereby confining the connectivity values to the range $[0, 1]$. We performed Pearson's correlation analysis to correlate original (c_O) and augmented (c_A) connectivity values with MMSE and age within OASIS-3 and ADNI-2 individually. Subsequently, we extended these analyses to examine correlations on the combined dataset before and after applying harmonization techniques, specifically the proposed distribution matching as well as ComBat with three covariate configurations (both sex and age, only sex, and neither). We then applied one-sided Wilcoxon signed-rank tests to the absolute values of the correlation coefficients between MMSE/age and structural brain connectivity to compare the effects of the harmonization methods against no harmonization. These tests were conducted once for each harmonization method against no harmonization. We used four measures to assess the performance: the absolute value of the correlation coefficient for each connection, the p-value generated by the abovementioned one-sided Wilcoxon signed-rank test, the number of brain connections surviving the Bonferroni correction (i.e., multiplication of the p-value by the number of connections, 3570), and the lowest Bonferroni-corrected p-value generated from the correlation analysis across brain connections.

2.5 Permutation within ADNI-2

To investigate the potential impact of the reduction in the number of subjects during internal harmonization, we conducted a permutation analysis focusing on the two brain connections most strongly correlated with age and MMSE. We selected a random subset of 106 subjects from the 209 subjects of the ADNI-2 dataset, mirroring the reduction in the number of subjects in our internal harmonization experiment within OASIS-3, and estimated the distribution of the selected connectivity value. We repeated this random selection 1000 times to create a robust permutation set, which helped to elucidate the potential variability introduced by sample size reduction relative to the gold-standard distribution obtained from the full set.

3 Results

3.1 Qualitative assessment of distribution matching

We measured how much the significance ($s := -\log p$) values, corresponding to the correlation between MMSE and structural brain connectivity, increase after harmonization, i.e. Δs, for all brain connections. We then sorted the connections with respect to their Δs and chose the five connections with the smallest, 25^{th} percentile, median, 75^{th} percentile, and largest Δs. As shown in Fig. 1, the peaks of the fitted gamma curves for ADNI-2 and OASIS-3 better align after (than before) harmonization.

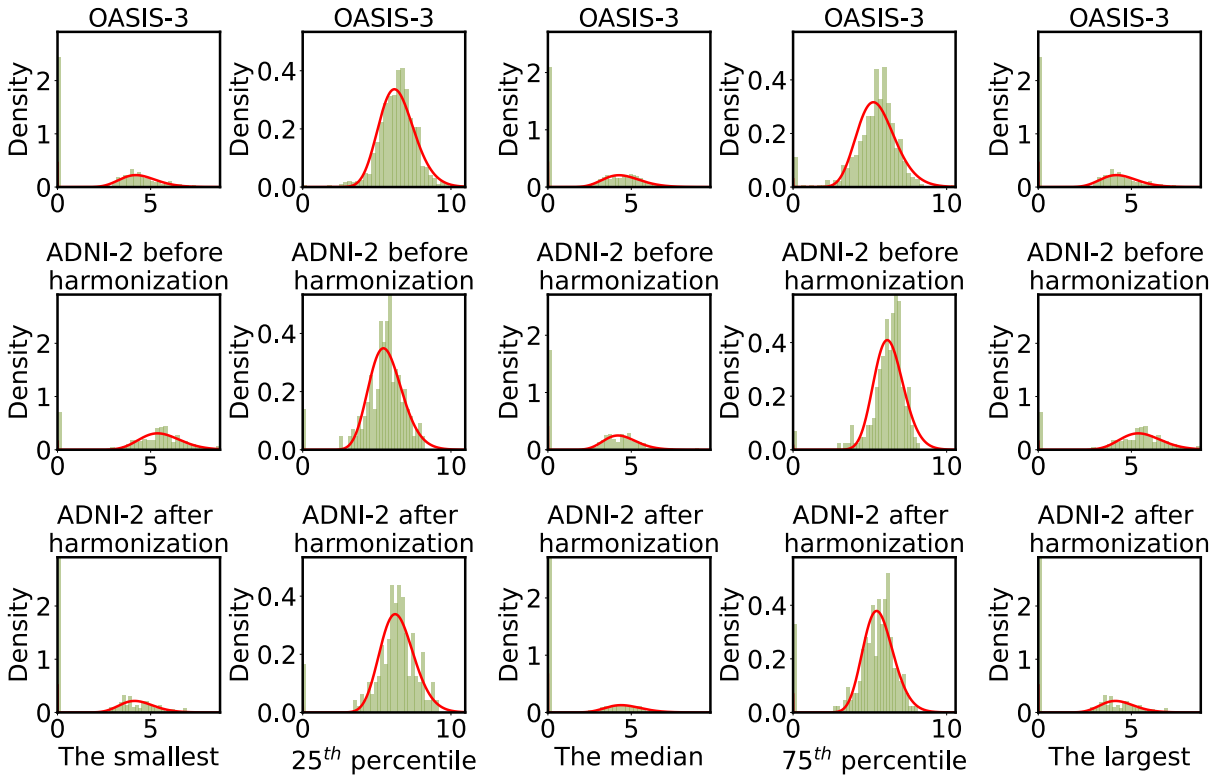

Figure 1: Normalized histograms of five connectivity values selected by choosing the smallest, 25$^{\text{th}}$ percentile, median, 75$^{\text{th}}$ percentile, and largest differences of significance values (Δs) from the correlation between MMSE and brain connectivity, before and after distribution-matching harmonization. Each column represents a structural brain connection, with the first row showing the reference OASIS-3 data and the second and third rows representing the ADNI-2 data before and after harmonization, respectively. The red curves show the fitted gamma distribution.

3.2 Correlation between harmonized structural connectivity and MMSE/age

We present the results of the quantitative evaluation of the proposed method in this section and in tables 2 and 3. Table 2 presents the correlations between age/MMSE and connectivity values, both within each dataset separately and in the combined dataset before and after harmonization. The comparison involves four harmonization approaches: the proposed distribution matching, ComBat with age and sex as covariates, ComBat with only sex as a covariate, and ComBat without any covariates. For the correlation between MMSE and original connectivity, distribution matching showed the highest mean absolute correlation (0.07 ± 0.05) and the highest number of significant relationships (361, also shown in Fig. 2(a)), with a minimum p-value of 1×10^{-16}. All three variations of ComBat showed slightly lower performance compared to distribution matching, with similar mean absolute correlations of 0.06 ± 0.05 and no more than 267 significant relationships. In correlations with augmented structural connectivity, distribution matching similarly outperformed other methods, showing the highest mean absolute correlation (0.17 ± 0.06) and the largest number of significant relationships (2406, shown in Fig. 2(b)), with a minimum p-value of 5×10^{-32}. This compares favorably to the no-harmonization case. ComBat harmonization incorporating both age and sex as covariates outperformed ComBat variants excluding age (both with and without sex as a covariate).

For MMSE correlations both with original and augmented structural connectivity, the distribution matching method had the most significant p-values from the Wilcoxon rank test. In the case of MMSE, we expected a positive correlation between MMSE and brain connectivity because lower MMSE indicates more advanced cognitive decline. Figure 3 (top row) demonstrates that distribution matching has enhanced the correlation between MMSE and (both original and augmented) structural connectivity, as visible in most of the dots being above the identity line.

For age correlations with original structural connectivity c_O, distribution matching and ComBat with age and sex as covariates showed a similar mean absolute correlation, while the former had the most significant p-value from one-sided Wilcoxon signed-rank test whereas the latter had the highest number of significant relationships and most significant minimum p-value. Both methods increased the mean absolute correlation and number of significant relationships after harmonization. ComBat variants excluding age (either with or without sex as a covariate) showed lower performance. In correlations with the augmented structural brain connectivity c_A, distribution matching demonstrated the best performance with a mean absolute correlation of 0.24 ± 0.08, the highest number of significant relationships (3228, see Fig. 2(d)), smallest p-values from one-sided Wilcoxon signed-rank test, and a minimum p-value of 8×10^{-60}. This was a significant improvement over the no-harmonization case. ComBat with age harmonization performed slightly worse, but still better than the two variants excluding age in covariates.

Table 2: The comparison of the correlations between age/MMSE and connectivity values before/after harmonization for the ADNI-2 and OASIS-3 datasets.

					[1]$\lvert r\rvert$ mean $\pm$ std	[2]One-sided Wilcoxon test, p_w	[3]$\#(p<0.05)$	[4]min p
MMSE	c_O	OASIS-3			0.06 ± 0.05	—	199	5×10^{-14}
		ADNI-2			0.07 ± 0.06	—	2	4×10^{-2}
		OASIS-3 and ADNI-2	No harmonization		0.06 ± 0.05	—	288	$\mathbf{2 \times 10^{-19}}$
			Distribution matching		$\mathbf{0.07 \pm 0.05}$	$\mathbf{8 \times 10^{-60}}$	**361**	1×10^{-16}
			ComBat	w/ age, w/ sex	0.06 ± 0.05	1	267	5×10^{-18}
				w/o age, w/ sex	0.06 ± 0.05	1	217	5×10^{-16}
				w/o age, w/o sex	0.06 ± 0.05	1	229	8×10^{-16}
	c_A	OASIS-3			0.15 ± 0.06	—	1634	6×10^{-25}
		ADNI-2			0.16 ± 0.09	—	288	6×10^{-6}
		OASIS-3 and ADNI-2	No harmonization		0.14 ± 0.07	—	1739	$\mathbf{6 \times 10^{-33}}$
			Distribution matching		$\mathbf{0.17 \pm 0.06}$	**0**	**2406**	5×10^{-32}
			ComBat	w/ age, w/ sex	0.16 ± 0.07	4×10^{-261}	2097	2×10^{-32}
				w/o age, w/ sex	0.15 ± 0.07	8×10^{-117}	1965	4×10^{-30}
				w/o age, w/o sex	0.15 ± 0.07	1×10^{-111}	1962	8×10^{-30}
Age	c_O	OASIS-3			0.11 ± 0.09	—	804	2×10^{-51}
		ADNI-2			0.09 ± 0.07	—	34	2×10^{-6}
		OASIS-3 and ADNI-2	No harmonization		0.10 ± 0.08	—	898	2×10^{-57}
			Distribution matching		$\mathbf{0.11 \pm 0.09}$	$\mathbf{4 \times 10^{-61}}$	1013	5×10^{-55}
			ComBat	w/ age, w/ sex	$\mathbf{0.11 \pm 0.09}$	0.001	**1074**	$\mathbf{1 \times 10^{-152}}$
				w/o age, w/ sex	0.09 ± 0.08	1	770	7×10^{-56}
				w/o age, w/o sex	0.10 ± 0.08	1	855	5×10^{-55}
	c_A	OASIS-3			0.24 ± 0.09	—	2874	5×10^{-50}
		ADNI-2			0.13 ± 0.09	—	181	7×10^{-8}
		OASIS-3 and ADNI-2	No harmonization		0.20 ± 0.09	—	2666	$\mathbf{5 \times 10^{-62}}$
			Distribution matching		$\mathbf{0.24 \pm 0.08}$	**0**	**3228**	8×10^{-60}
			ComBat	w/ age, w/ sex	0.22 ± 0.08	4×10^{-232}	2970	9×10^{-60}
				w/o age, w/ sex	0.21 ± 0.08	8×10^{-94}	2895	4×10^{-55}
				w/o age, w/o sex	0.21 ± 0.08	1×10^{-88}	2901	1×10^{-54}

[1] Mean and standard deviation of the absolute values of the correlation coefficients.
[2] p-value generated from one-sided Wilcoxon signed-rank test, comparing r with the no-harmonization case.
[3] Number of significant relationships with p-values surviving the Bonferroni correction.
[4] Minimum Bonferroni-corrected p-value of Pearson's correlation.

Table 3: The comparison of the correlations between age/MMSE and connectivity values before/after harmonization for internal harmonization within OASIS-3 dataset.

				[1]$\lvert r\rvert$ mean $\pm$ std	[2]One-sided Wilcoxon test, p_w	[3]$\#(p < 0.05)$	[4]min p
MMSE	c_O	No harmonization		$\mathbf{0.06 \pm 0.05}$	—	202	3×10^{-13}
		Distribution matching		$\mathbf{0.06 \pm 0.05}$	1	182	5×10^{-14}
		ComBat	w/ age, w/ sex	$\mathbf{0.06 \pm 0.05}$	1×10^{-8}	**207**	$\mathbf{2 \times 10^{-14}}$
			w/o age, w/ sex	$\mathbf{0.06 \pm 0.05}$	2×10^{-18}	202	$\mathbf{2 \times 10^{-14}}$
			w/o age, w/o sex	$\mathbf{0.06 \pm 0.05}$	$\mathbf{4 \times 10^{-41}}$	203	$\mathbf{2 \times 10^{-14}}$
	c_A	No harmonization		0.15 ± 0.06	—	1642	2×10^{-24}
		Distribution matching		0.15 ± 0.07	1	1612	8×10^{-24}
		ComBat	w/ age, w/ sex	0.15 ± 0.06	2×10^{-290}	1708	3×10^{-25}
			w/o age, w/ sex	$\mathbf{0.16 \pm 0.07}$	**0**	1787	2×10^{-25}
			w/o age, w/o sex	$\mathbf{0.16 \pm 0.07}$	**0**	**1792**	$\mathbf{1 \times 10^{-25}}$
Age	c_O	No harmonization		$\mathbf{0.11 \pm 0.09}$	—	805	5×10^{-51}
		Distribution matching		$\mathbf{0.11 \pm 0.09}$	1	795	8×10^{-50}
		ComBat	w/ age, w/ sex	$\mathbf{0.11 \pm 0.09}$	1	**879**	$\mathbf{1 \times 10^{-94}}$
			w/o age, w/ sex	0.10 ± 0.09	1	801	1×10^{-49}
			w/o age, w/o sex	0.10 ± 0.09	1	800	1×10^{-49}
	c_A	No harmonization		$\mathbf{0.24 \pm 0.09}$	—	2864	$\mathbf{7 \times 10^{-50}}$
		Distribution matching		$\mathbf{0.24 \pm 0.09}$	$\mathbf{4 \times 10^{-154}}$	**2921**	1×10^{-49}
		ComBat	w/ age, w/ sex	0.22 ± 0.09	1	2766	2×10^{-47}
			w/o age, w/ sex	0.23 ± 0.09	1	2812	3×10^{-49}
			w/o age, w/o sex	0.23 ± 0.09	1	2812	3×10^{-49}

[1] Mean and standard deviation of the absolute values of the correlation coefficients.
[2] p-value generated from one-sided Wilcoxon signed-rank test, comparing r with the no-harmonization case.
[3] Number of significant relationships with p-values surviving the Bonferroni correction.
[4] Minimum Bonferroni-corrected p-value of Pearson's correlation.

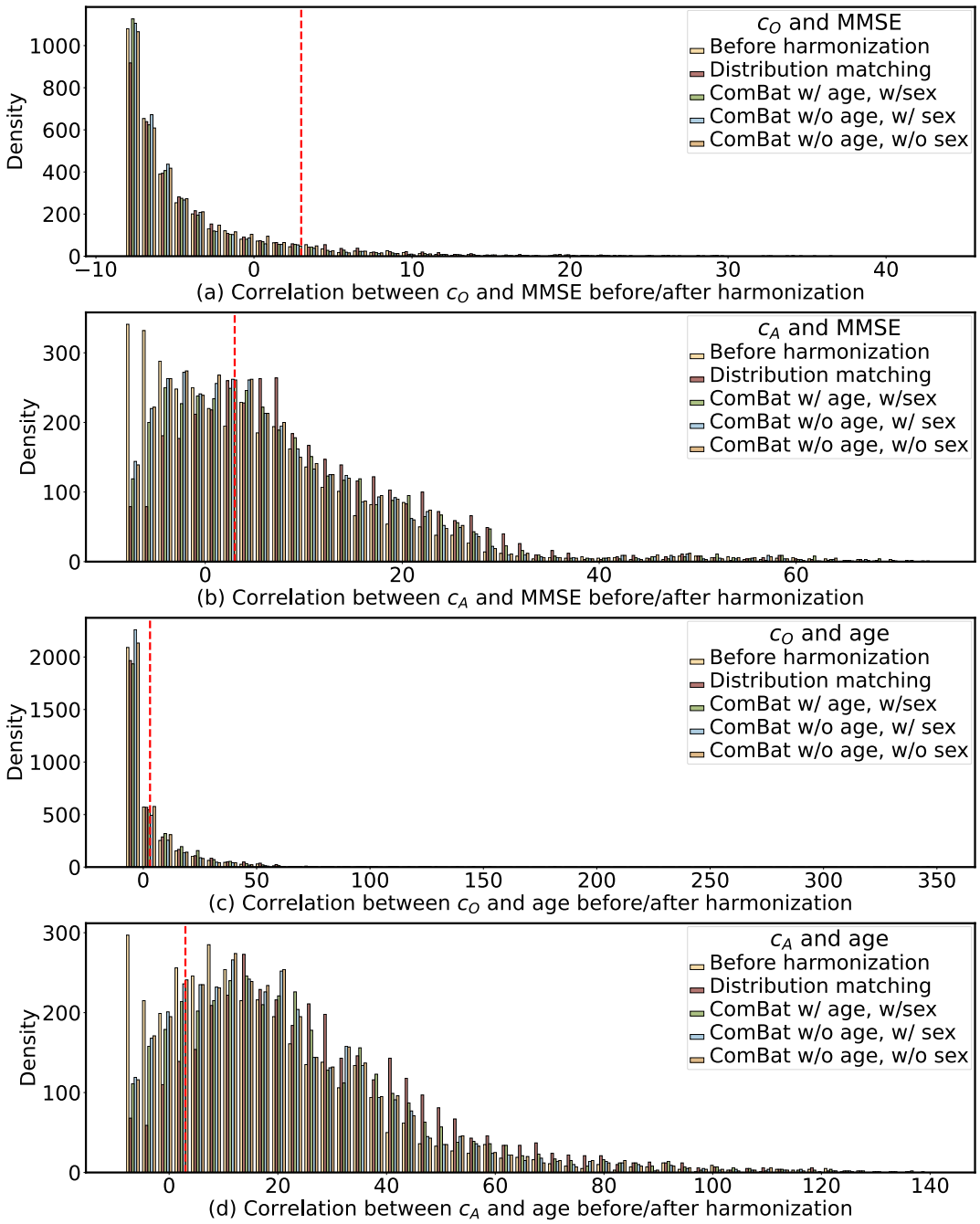

Figure 2: The histogram of $s := -\log p$ from Pearson's correlation between original (a,c) or augmented (b,d) structural brain connectivity and MMSE (a,b) or age (c,d), before and after harmonization. The red dashed vertical line represents the significance cutoff threshold of $-\log 0.05$.

In the case of age, we expected a negative correlation between age and structural brain connectivity. This is well illustrated in Fig. 3 (bottom), where distribution matching has improved the correlation between age and (both original and augmented) structural connectivity, as evidenced by most of the dots being

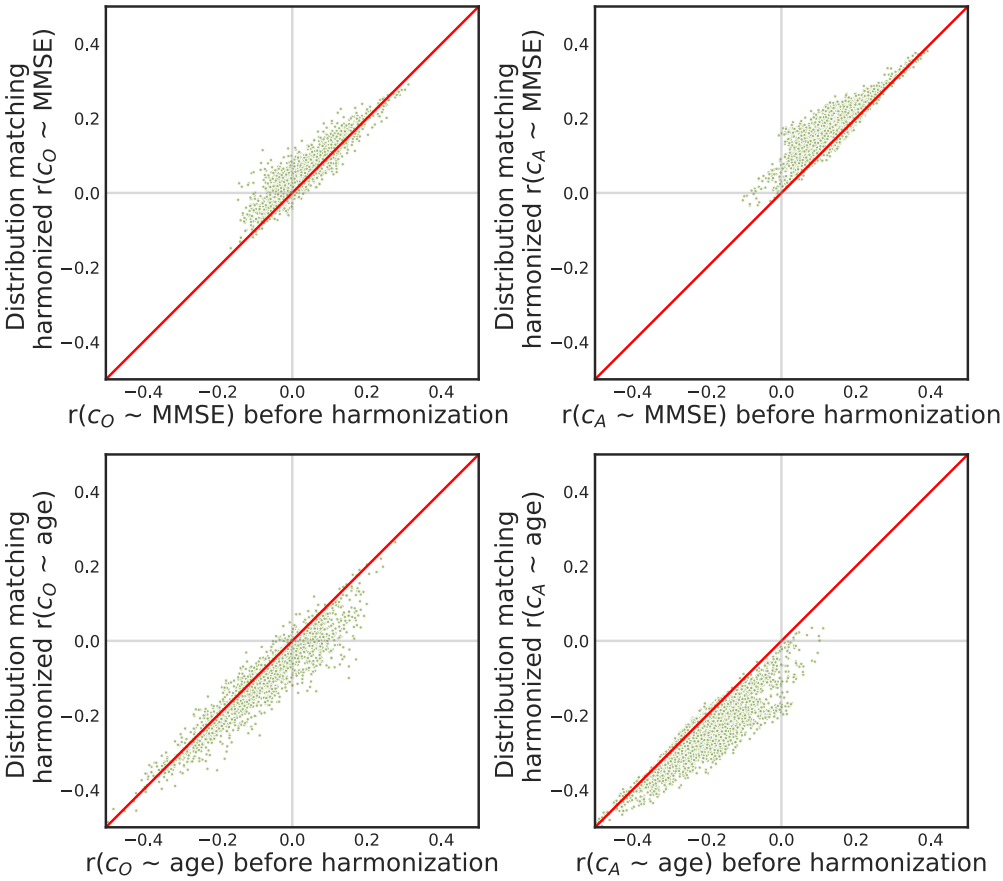

Figure 3: Scatter plots comparing post- vs. pre-harmonization Pearson's correlation (r) between MMSE (top) or age (bottom) and original (left) or augmented (right) structural brain connectivity. The identity line is plotted in red. In all cases, the proposed distribution-matching harmonization increased the overall $|r|$.

below the identity line.

Next, since the OASIS-3 dataset includes data from two types of Siemens MRI scanners, we performed internal harmonization within that dataset, the results of which are shown in Table 3. Different harmonization methods resulted in only small differences in the mean absolute correlations and the number of significant relationships. In particular, for correlations with age, no significant improvements were observed in the mean absolute correlations except for the correlation with augmented connectivity after distribution matching ($p = 4 \times 10^{-154}$). For MMSE correlations with original and augmented structural connectivity, ComBat methods slightly improved the mean absolute correlations, whereas distribution matching showed no improvement.

Lastly, regarding the permutation experiments within ADNI-2 (Section 2.5), the PDFs of the 1000 permutations of the subsampled datasets and the full dataset are visualized in Figure 4, illustrating the

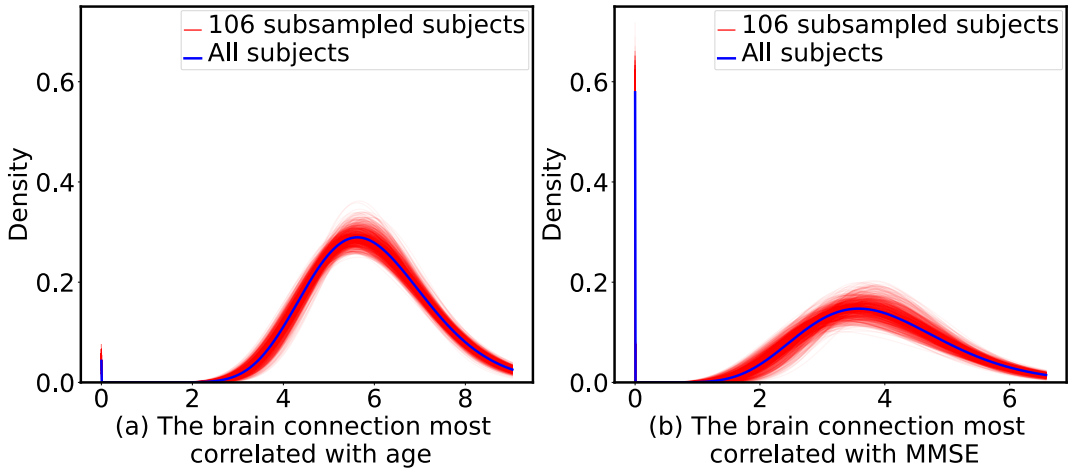

Figure 4: Distribution of the PDFs of the brain connection most correlated with (a) age and (b) MMSE. Each plot compares 1000 random permutations to select a subset of 106 subjects (red curves), to the full dataset with 209 subjects (blue curve), in ADNI-2. The brain connection most correlated with age is between the left thalamus and left hippocampus. The connection most correlated with MMSE is between the left lingual cortex and left middle temporal cortex.

extent of variability introduced by cohort size reduction.

4 Discussion

In this study, we proposed a distribution-matching technique for harmonizing multi-site structural brain connectivity data, and evaluated it while comparing it with the ComBat harmonization approach. We analyzed structural brain connectivity data from two datasets, OASIS-3 and ADNI-2, and examined the impact of harmonization on correlations between brain connectivity and both MMSE and age. We evaluated the efficacy of our distribution matching method both qualitatively and quantitatively, with both assessments showing robust performance. Qualitatively, our distribution matching harmonization achieved the expected distributional alignment of ADNI-2 to OASIS-3. The effectiveness of harmonization was assessed quantitatively by comparing correlation strengths and statistical significance after applying each harmonization method to those before harmonization. Consistently through our cross-dataset analyses, we found distribution matching almost always produced higher mean absolute correlations and more significant p-values compared to ComBat (as shown in Fig. 2 and Table 2), suggesting it is a more effective method for harmonizing structural connectivity data in these datasets. Our findings provide insight into the complexities of harmonizing multi-site structural brain connectivity data and lay the groundwork for further refinements in harmonization techniques.

While minimum p-values can provide insight into the strongest individual correlations, relying solely on this measure has some limitations in the context of large-scale neuroimaging studies. In datasets involving thousands of brain connections, extremely low p-values can occur by chance, potentially leading to misleading conclusions about harmonization effectiveness. Moreover, minimum p-values are not robust to outliers or extreme values, which could disproportionately affect the final conclusions. To address these limitations and provide a more comprehensive assessment of harmonization performance, we incorporated the following additional measures. Absolute correlation coefficients offer a direct indication of the strength of association between brain connectivity and MMSE/age, providing insight into effect sizes that are theoretically independent of sample size and reflect practical significance. Wilcoxon signed-rank tests, comparing results before and after applying different harmonization methods, help us understand whether harmonization significantly improves the overall strength of correlations across all connections, providing a whole-brain measure of effectiveness. (This test requires the conditional distribution of each observation given the others to be symmetric about a common point, which may, however, not always be satisfied in our case, hence reduced accuracy of the Wilcoxon p-values.) The number of connections surviving Bonferroni correction reflects the overall sensitivity for detecting meaningful correlations while controlling for multiple comparisons, providing insight into the breadth of harmonization effects throughout all brain connections. By utilizing these multiple measures, we aim to capture different aspects of harmonization performance. This offers a more comprehensive and nuanced evaluation that enables the assessment of both the strength and prevalence of brain-behavior relationships across the entire brain, providing a more reliable basis for comparing different harmonization techniques.

Our evaluation consisted of distribution matching harmonization on the combined OASIS-3 and ADNI-2 datasets, as well as internal harmonization within the OASIS-3 dataset. Interestingly, we observed that internal harmonization within OASIS-3 rarely improved the correlation results, regardless of whether our distribution matching or the ComBat method was used (see Table 3). This outcome can be attributed to the following two factors. The OASIS-3 data has a high degree of consistency and standardization, e.g. in terms of data collection protocols, acquisition parameters, and possibly some preprocessing. If there is minimal variability between different scanner types or parameters within OASIS-3, the impact of harmonization will be negligible, thereby making internal harmonization less beneficial. Furthermore, in the harmonization of the combined OASIS-3 and ADNI-2 datasets, we used OASIS-3 as the reference site, which included 761 subjects, while the ADNI-2 dataset with 209 subjects served as the new site. For the internal harmonization within OASIS-3, we divided OASIS-3 into two sites: one with 655 subjects (used as the reference site) and the other with 106 subjects (used as the new site). The reduced number of subjects during internal harmonization (especially the number of subjects in the new site decreasing from 209 to 106), could be a contributing factor. We performed permutation tests by randomly selecting 106

subjects from the total 209 subjects in the ADNI-2 dataset to simulate the reduction in the sample size in our internal harmonization experiment, and computed the distribution of the selected feature. As shown in Figure 4, we found a non-negligible variation among the permuted PDFs (red curves), which often deviated from the full sample distribution (blue curve) to some extent. This suggests that a smaller new site might represent the variability and distribution less precisely, leading to less robust harmonization. The lack of improvement in correlation results within the OASIS-3 dataset after internal harmonization could therefore be attributed to the possible homogeneity of the dataset and the smaller new site in internal harmonization. Accounting for these factors can help us refine harmonization techniques and better tailor them to specific datasets, ultimately improving the reliability and consistency of large-scale multi-site neuroimaging analyses.

We compared the distribution matching method with three variants of ComBat harmonization on structural connectivity data: ComBat with both age and sex as covariates, ComBat with only sex as a covariate, and ComBat without covariates. The inclusion of age or sex as covariates aims to remove site-specific variability attributable to these demographic factors, potentially harmonizing data more effectively across different sites, age ranges, and sex groups (Pomponio et al., 2020; Zhou et al., 2023). However, this approach results in harmonized connectivity values that are partially conditioned on age and/or sex. We observed significant correlations between sex, age, and MMSE across all subjects in our study (sex with age, $p = 9 \times 10^{-6}$; age with MMSE, $p = 3 \times 10^{-18}$; sex with MMSE, $p = 3 \times 10^{-5}$). These relationships raise concerns about potential information leakage when including age and/or sex as covariates during harmonization, particularly when subsequent analyses involve correlating these very variables with harmonized structural connectivity. As shown in Table 2, ComBat with age and sex as covariates generally produced better correlation coefficients than ComBat without them. While this suggests improved effectiveness of harmonization, it is crucial to consider the potential for information leakage. The enhanced correlations may partly result from age and sex influences being embedded in the harmonized values, potentially biasing the results and even creating spurious correlations. Notably, our distribution matching harmonization method did not incorporate age or sex information, thanks to which it avoided the risk of age- or sex-related information leakage [1]. By excluding these covariates, distribution matching may offer a more straightforward and potentially less biased evaluation of how structural brain connectivity relates to age, sex, and other variables of interest.

As shown in Fig. 1, we selected five brain connections by ranking the difference of negative logarithm of p-values (Δs) generated from the correlation between MMSE and structural brain connectivity before vs. after harmonization, and choosing the smallest, 25^{th} percentile, median, 75^{th} percentile, and largest ones.

[1] We could have controlled for sex in our distribution-matching approach simply by applying it to the two sex groups separately.

We chose these five examples in this way so they were selected in an unbiased manner. This, however, might create the expectation that the harmonization-caused shift in the peak of the histogram would increase with respect to Δs. However, as shown in the figure, the amount of change after harmonization does not necessarily correlate with the ranking of Δs. It is worth noting that Δs is related to not only the harmonization effect size, but likely also to the significance (s) of the correlation with MMSE itself, with the latter varying largely across different structural brain connections and not expected to be necessarily related to the effectiveness of harmonization. In other words, brain connections that benefit the most from harmonization and possibly go through a sizable distribution shift may not necessarily coincide with those that are highly correlated with MMSE and have a high Δs.

Our current study has several limitations that present opportunities for future research. First, our analysis was confined to two imaging datasets and one tractography method, which limits the generalizability of our findings. Future studies should extend this work to include multiple datasets from diverse sites and data processed with various tractography methods, allowing for a more comprehensive evaluation of the distribution matching method. Second, to further validate the effectiveness of our harmonization approach, downstream analyses, such as predictive modeling, classification, and more complex network analyses, could be employed. Such analyses would provide concrete evidence of the impact of harmonization on real-world applications in neuroimaging research, potentially revealing subtle effects that might be obscured by harmonization methods less tailored to structural brain connectivity. Finally, future work could investigate the impact of harmonization on longitudinal data, explore its effectiveness in harmonizing different imaging modalities, and examine its performance in clinical populations with specific neurological or psychiatric conditions. Addressing these limitations and exploring these new directions, which would enhance the robustness, applicability, and sensitivity of our distribution-matching harmonization in multi-site neuroimaging studies, is part of our ongoing research.

Data and Code Availability

Data used in this study were obtained from two public databases: the third phase of the Open Access Series of Imaging Studies 3 (OASIS-3) (LaMontagne et al., 2019), which is freely available from www.oasis-brains.org, and the second phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI-2), which is available for download from http://adni.loni.usc.edu for researchers who meet the criteria for access to these data.

The code for reconstructing the diffusion orientation distribution function in constant solid angle, performing Hough-transform global probabilistic tractography, computing the connectivity matrix, and augment-

ing it with indirect connections is available at www.nitrc.org/projects/csaodf-hough. Combat data harmonization was performed using a package available at https://github.com/rpomponio/neuroHarmonize. FreeSurfer (Fischl, 2012) and FSL (Jenkinson et al., 2012) are available to download at https://freesurfer.net and https://fsl.fmrib.ox.ac.uk, respectively.

We are currently making our distribution-matching codes user-friendly and will deposit them in an open-access platform, providing the link here upon the acceptance of this manuscript.

Author Contributions

Zhen Zhou: Conceptualization, Methodology, Formal analysis, Validation, Writing - original draft, and Writing - review & editing. Bruce Fischl: Conceptualization, Methodology, and Writing - review & editing. Iman Aganj: Data curation, Conceptualization, Methodology, Formal analysis, Writing - review & editing, and Funding acquisition.

Funding

Support for this research was provided by the Michael J. Fox Foundation for Parkinson's Research (MRI Biomarkers Program award MJFF-021226).

Additional support was provided by the National Institutes of Health (NIH), specifically in part by the National Institute on Aging (RF1AG068261, R01AG068261, R21AG082082, R01AG064027, R01AG008122, R01AG016495, R01AG070988), the BRAIN Initiative Cell Census Network grants U01MH117023 and UM1MH130981, the Brain Initiative Brain Connects consortium (U01NS132181, UM1NS132358), the National Institute for Biomedical Imaging and Bioengineering (R01EB023281, R01EB006758, R21EB018907, R01EB019956, P41EB030006), the National Institute of Mental Health (UM1MH130981, R01MH123195, R01MH121885, RF1MH123195), the National Institute for Neurological Disorders and Stroke (R01NS0525851, R21NS072652, R01NS070963, R01NS083534, U01NS086625, U24NS10059103, R01NS105820), and was made possible by the resources provided by Shared Instrumentation Grants S10RR023401, S10RR019307, and S10RR023043. Additional support was provided by the NIH Blueprint for Neuroscience Research (U01MH093765), part of the multi-institutional Human Connectome Project.

Data collection and sharing for this project was funded in part by OASIS-3: Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50

AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352.

Data were also provided in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (NIH Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the NIA, the NIBIB, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the NIH (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Declaration of Competing Interests

B. Fischl is an advisor to DeepHealth, a company whose medical pursuits focus on medical imaging and measurement technologies. His interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict-of-interest policies. The other authors have nothing to disclose.

Acknowledgments

We thank Dr. Aina Frau-Pascual for previously helping us in data curation.

References

Aganj, I., Lenglet, C., Jahanshad, N., Yacoub, E., Harel, N., Thompson, P. M., & Sapiro, G. (2011). A hough transform global probabilistic approach to multiple-subject diffusion MRI tractography. *Medical image analysis*, *15*(4), 414–425. https://doi.org/10.1016/j.media.2011.01.003

Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., & Harel, N. (2010). Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magnetic resonance in medicine*, *64*(2), 554–566. https://doi.org/10.1002/mrm.22365

Aganj, I., Mora, J., Frau-Pascual, A., Fischl, B., & Initiative, A. D. N. (2023). Exploratory correlation of the human structural connectome with non-MRI variables in alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *15*(4), e12511. https://doi.org/10.1002/dad2.12511

Aganj, I., Prasad, G., Srinivasan, P., Yendiki, A., Thompson, P. M., & Fischl, B. (2014). Structural brain network augmentation via kirchhoff's laws. *Joint Annual Meeting of ISMRM-ESMRMB, http://nmr. mgh. harvard. edu/˜ iman/ConductanceModel_ISMRM14_iman. pdf*, *22*, 2665.

Bazinet, V., Hansen, J. Y., & Misic, B. (2023). Towards a biologically annotated brain connectome. *Nature reviews neuroscience*, *24*(12), 747–760. https://doi.org/10.1038/s41583-023-00752-3

Beck, D., de Lange, A.-M. G., Maximov, I. I., Richard, G., Andreassen, O. A., Nordvik, J. E., & Westlye, L. T. (2021). White matter microstructure across the adult lifespan: A mixed longitudinal and cross-sectional study using advanced diffusion models and brain-age prediction. *NeuroImage*, *224*, 117441. https://doi.org/10.1016/j.neuroimage.2020.117441

Beckett, L. A., Donohue, M. C., Wang, C., Aisen, P., Harvey, D. J., Saito, N., Initiative, A. D. N., et al. (2015). The Alzheimer's disease neuroimaging initiative phase 2: Increasing the length, breadth, and depth of our understanding. *Alzheimer's & Dementia*, *11*(7), 823–831. https://doi.org/10.1016/j.jalz.2015.05.004

Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *neuroimage*, *34*(1), 144–155. https://doi.org/10.1016/j.neuroimage.2006.09.018

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

De Luca, A., Karayumak, S. C., Leemans, A., Rathi, Y., Swinnen, S., Gooijers, J., Clauwaert, A., Bahr, R., Sandmo, S. B., Sochen, N., et al. (2022). Cross-site harmonization of multi-shell diffusion MRI measures based on rotational invariant spherical harmonics (RISH). *NeuroImage*, *259*, 119439. https://doi.org/10.1016/j.neuroimage.2022.119439

Evans, M. J., & Rosenthal, J. S. (2004). *Probability and statistics: The science of uncertainty*. Macmillan.

Fischl, B. (2012). Freesurfer. *Neuroimage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., et al. (2018). Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*, *167*, 104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024

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

Frau-Pascual, A., Augustinack, J., Varadarajan, D., Yendiki, A., Salat, D. H., Fischl, B., Aganj, I., & Initiative, A. D. N. (2021). Conductance-based structural brain connectivity in aging and dementia. *Brain connectivity*, *11*(7), 566–583. https://doi.org/10.1089/brain.2020.0903

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Huynh, K. M., Chen, G., Wu, Y., Shen, D., & Yap, P.-T. (2019). Multi-site harmonization of diffusion MRI data via method of moments. *IEEE transactions on medical imaging*, *38*(7), 1599–1609. https://doi.org/10.1109/tmi.2019.2895020

Jack Jr, C. R., Barnes, J., Bernstein, M. A., Borowski, B. J., Brewer, J., Clegg, S., Dale, A. M., Carmichael, O., Ching, C., DeCarli, C., et al. (2015). Magnetic resonance imaging in alzheimer's disease neuroimaging initiative 2. *Alzheimer's & Dementia*, *11*(7), 740–756. https://doi.org/10.1016/j.jalz.2015.05.002

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

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

Koppers, S., Bloy, L., Berman, J. I., Tax, C. M., Edgar, J. C., & Merhof, D. (2019). Spherical harmonic residual network for diffusion signal harmonization. *Computational Diffusion MRI: International MICCAI Workshop, Granada, Spain, September 2018 22*, 173–182. https://doi.org/10.1007/978-3-030-05831-9_14

Koppers, S., Haarburger, C., & Merhof, D. (2017). Diffusion MRI signal augmentation: From single shell to multi shell with deep learning. *Computational Diffusion MRI: MICCAI Workshop, Athens, Greece, October 2016 19*, 61–70. https://doi.org/10.1007/978-3-319-54130-3_5

Kurokawa, R., Kamiya, K., Koike, S., Nakaya, M., Uematsu, A., Tanaka, S. C., Kamagata, K., Okada, N., Morita, K., Kasai, K., et al. (2021). Cross-scanner reproducibility and harmonization of a diffusion mri structural brain network: A traveling subject study of multi-b acquisition. *NeuroImage*, *245*, 118675. https://doi.org/10.1016/j.neuroimage.2021.118675

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., et al. (2019). OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *MedRxiv*, 2019–12. https://doi.org/10.1101/2019.12.13.19014902

Magnotta, V. A., Matsui, J. T., Liu, D., Johnson, H. J., Long, J. D., Bolster Jr, B. D., Mueller, B. A., Lim, K., Mori, S., Helmer, K. G., et al. (2012). Multicenter reliability of diffusion tensor imaging. *Brain connectivity*, *2*(6), 345–355. https://doi.org/10.1089/brain.2012.0112

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al. (2011). The Parkinson progression marker initiative (PPMI). *Progress in neurobiology*, *95*(4), 629–635. https://doi.org/10.1016/j.pneurobio.2011.09.005

Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Grant, G., Marx, C. E., Morey, R. A., Flashman, L. A., et al. (2016). Inter-site and inter-scanner diffusion MRI data harmonization. *NeuroImage*, *135*, 311–323. https://doi.org/10.1016/j.neuroimage.2016.04.041

Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *45*(2), 265–269. https://doi.org/10.1002/1531-8249(199902)45:2%3C265::AID-ANA21%3E3.0.CO;2-3

Moyer, D., Ver Steeg, G., Tax, C. M., & Thompson, P. M. (2020). Scanner invariant representations for diffusion MRI harmonization. *Magnetic resonance in medicine*, *84*(4), 2174–2189. https://doi.org/10.1002/mrm.28243

Onicas, A. I., Ware, A. L., Harris, A. D., Beauchamp, M. H., Beaulieu, C., Craig, W., Doan, Q., Freedman, S. B., Goodyear, B. G., Zemek, R., et al. (2022). Multisite harmonization of structural DTI networks in children: An A-CAP study. *Frontiers in Neurology*, *13*, 850642. https://doi.org/10.3389/fneur.2022.850642

Panman, J. L., To, Y. Y., van der Ende, E. L., Poos, J. M., Jiskoot, L. C., Meeter, L. H., Dopper, E. G., Bouts, M. J., van Osch, M. J., Rombouts, S. A., et al. (2019). Bias introduced by multiple head coils in mri research: An 8 channel and 32 channel coil comparison. *Frontiers in neuroscience*, *13*, 729. https://doi.org/10.3389/fnins.2019.00729

Patel, J., Schöttner, M., Tarun, A., Tourbier, S., Alemán-Gómez, Y., Hagmann, P., & Bolton, T. A. (2024). Modeling the impact of mri acquisition bias on structural connectomes: Harmonizing structural connectomes. *Network Neuroscience*, 1–30. https://doi.org/10.1162/netn_a_00368

Pines, A. R., Cieslak, M., Larsen, B., Baum, G. L., Cook, P. A., Adebimpe, A., Dávila, D. G., Elliott, M. A., Jirsaraie, R., Murtha, K., et al. (2020). Leveraging multi-shell diffusion for studies of brain development in youth and young adulthood. *Developmental cognitive neuroscience*, *43*, 100788. https://doi.org/10.1016/j.dcn.2020.100788

Pinto, M. S., Paolella, R., Billiet, T., Van Dyck, P., Guns, P.-J., Jeurissen, B., Ribbens, A., den Dekker, A. J., & Sijbers, J. (2020). Harmonization of brain diffusion MRI: Concepts and methods. *Frontiers in Neuroscience*, *14*, 396. https://doi.org/10.3389/fnins.2020.00396

Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I. M., Satterthwaite, T. D., Fan, Y., et al. (2020). Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, *208*, 116450. https://doi.org/10.1016/j.neuroimage.2019.116450

Sotiropoulos, S. N., & Zalesky, A. (2019). Building connectomes using diffusion mri: Why, how and but. *NMR in Biomedicine*, *32*(4), e3752. https://doi.org/10.1002/nbm.3752

Tax, C. M., Grussu, F., Kaden, E., Ning, L., Rudrapatna, U., Evans, C. J., St-Jean, S., Leemans, A., Koppers, S., Merhof, D., et al. (2019). Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms. *NeuroImage*, *195*, 285–299. https://doi.org/10.1016/j.neuroimage.2019.01.077

Tournier, J.-D., Mori, S., & Leemans, A. (2011). Diffusion tensor imaging and beyond. *Magnetic resonance in medicine*, *65*(6), 1532. https://doi.org/10.1002/mrm.22924

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Donohue, M. C., Green, R. C., Harvey, D., Jack Jr, C. R., et al. (2015). Impact of the alzheimer's disease neuroimaging initiative, 2004 to 2014. *Alzheimer's & Dementia*, *11*(7), 865–884. https://doi.org/10.1016/j.jalz.2015.04.005

Wheeler, A. L., & Voineskos, A. N. (2014). A review of structural neuroimaging in schizophrenia: From connectivity to connectomics. *Frontiers in human neuroscience*, *8*, 653. https://doi.org/10.3389/fnhum.2014.00653

Wrobel, J., Martin, M., Bakshi, R., Calabresi, P. A., Elliot, M., Roalf, D., Gur, R. C., Gur, R. E., Henry, R. G., Nair, G., et al. (2020). Intensity warping for multisite MRI harmonization. *NeuroImage*, *223*, 117242. https://doi.org/10.1016/j.neuroimage.2020.117242

Xia, Y., & Shi, Y. (2022). Personalized DMRI harmonization on cortical surface. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 717–725. https://doi.org/10.1007/978-3-031-16446-0_68

Xia, Y., & Shi, Y. (2024). Diffusion MRI harmonization via personalized template mapping. *Human Brain Mapping*, *45*(5), e26661. https://doi.org/10.1002/hbm.26661

Yeh, C.-H., Jones, D. K., Liang, X., Descoteaux, M., & Connelly, A. (2021). Mapping structural connectivity using diffusion MRI: Challenges and opportunities. *Journal of Magnetic Resonance Imaging*, *53*(6), 1666–1682. https://doi.org/10.1002/jmri.27188

Yu, M., Linn, K. A., Cook, P. A., Phillips, M. L., McInnis, M., Fava, M., Trivedi, M. H., Weissman, M. M., Shinohara, R. T., & Sheline, Y. I. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. *Human brain mapping*, *39*(11), 4213–4227. https://doi.org/10.1002/hbm.24241

Zhang, R., Oliver, L. D., Voineskos, A. N., & Park, J. Y. (2023). RELIEF: A structured multivariate approach for removal of latent inter-scanner effects. *Imaging Neuroscience*, *1*, 1–16. https://doi.org/10.1162/imag_a_00011

Zhou, Z., Li, H., Srinivasan, D., Abdulkadir, A., Nasrallah, I. M., Wen, J., Doshi, J., Erus, G., Mamourian, E., Bryan, N. R., et al. (2023). Multiscale functional connectivity patterns of the aging brain learned from harmonized rsfMRI data of the multi-cohort iSTAGING study. *Neuroimage*, *269*, 119911. https://doi.org/10.1016/j.neuroimage.2023.119911

Zhou, Z., Srinivasan, D., Li, H., Abdulkadir, A., Shou, H., Davatzikos, C., Fan, Y., Consortium, I., et al. (2022). Harmonization of multi-site functional connectivity measures in tangent space improves brain age prediction. *Proceedings of SPIE–the International Society for Optical Engineering*, *12036*. https://doi.org/10.1117/12.2611557

Zhu, A. H., Moyer, D. C., Nir, T. M., Thompson, P. M., & Jahanshad, N. (2019). Challenges and opportunities in dMRI data harmonization. *Computational Diffusion MRI: International MICCAI Workshop, Granada, Spain, September 2018 22*, 157–172. https://doi.org/10.1007/978-3-030-05831-9_13