

Number of words in the abstract: 290
 Number of words in the main text: 2400
 Number of figures: 3
 Number of tables: 1
 Supplements: 2 figures

Exploring the limits of ComBat method for multi-site diffusion MRI harmonization

Suheyly Cetin-Karayumak, Ph.D.^a, Katharina Stegmayer, MD^c, Sebastian Walther, MD^c, Philip R. Szeszko, Ph.D.^d, Tim Crow, MD, PhD^e, Anthony James, MD^e, Matcheri Keshavan, MD^f, Marek Kubicki, MD, Ph.D.^a, Yogesh Rathi, Ph.D.^{a,b}

- ^a Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.
- ^b Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.
- ^c University of Bern, University Hospital of Psychiatry, Switzerland.
- ^d Mental Illness Research, Education and Clinical Center, James J. Peters VA Medical Center, Bronx, NY, USA; Department of Psychiatry, Icahn School of Medicine at Mount Sinai, NY, NY, USA.
- ^e Department of Psychiatry, SANE POWIC, Warneford Hospital, University of Oxford, UK
- ^f Department of Psychiatry, Beth Israel Deaconess Medical Centre, Harvard Medical School, Boston, Massachusetts, USA

Correspondence:

Suheyly Cetin-Karayumak, Psychiatry Neuroimaging Laboratory
 Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School
 2nd floor Room 264, 1249 Boylston St, 02215, Boston, Massachusetts, USA.
 t: +1 617-5256207
 e: skarayumak@bwh.harvard.edu

Abstract:

The findings from diffusion-weighted magnetic resonance imaging (dMRI) studies often show inconsistent and sometimes contradictory results due to small sample sizes as well as differences in acquisition parameters and pre-/post-processing methods. To address these challenges, collaborative multi-site initiatives have provided an opportunity to collect larger and more diverse groups of subjects, including those with neuropsychiatric disorders, leading to increased power and findings that may be more representative at the group and individual level. With the availability of these datasets openly, the ability of joint analysis of multi-site dMRI data has become more important than ever. However, intrinsic- or acquisition-related variability in scanner models, acquisition protocols, and reconstruction settings hinder pooling multi-site dMRI directly. One powerful and fast statistical harmonization method called ComBat (<https://github.com/Jfortin1/ComBatHarmonization>) was developed to mitigate the “batch effect” in gene expression microarray data and was adapted for multi-site dMRI harmonization to reduce scanner/site effect. Our goal is to evaluate this commonly used harmonization approach using a large diffusion MRI dataset involving 542 individuals from 5 sites. We investigated two important aspects of using ComBat for harmonization of fractional anisotropy (FA) across sites: First, we assessed how well ComBat preserves the inter-subject biological variability (measured by the effect sizes of between-group FA differences) after harmonization. Second, we evaluated the effect of minor differences in pre-processing on ComBat’s performance. While the majority of effect sizes are mostly preserved in some sites after harmonization, they are not well-preserved at other sites where non-linear scanner contributions exist. Further, even minor differences in pre-processing can yield unwanted effects during ComBat harmonization. Thus, our findings suggest paying careful attention to the data being harmonized as well as using the same processing pipeline while using ComBat for data harmonization.

Introduction:

Diffusion-weighted magnetic resonance imaging (dMRI) is the only non-invasive method that can map the living human brain's structural connections. The measures derived from dMRI are quite sensitive to the subtle changes of the underlying microstructural tissue and have contributed to a wealth of knowledge about the abnormalities in several neurological and psychiatric disorders^{1,2}. However, results from neuroimaging studies often show inconsistent and sometimes contradictory results due to small sample sizes as well as differences in acquisition parameters and processing methods^{3,4}. Large collaborative efforts have led to Big-Data initiatives aimed at improving the poor reproducibility of small neuroimaging studies and gaining statistical power to detect subtle changes in neuropsychiatric disorders⁵⁻⁹. However, scanner induced bias in dMRI signal limits direct pooling of the multi-site dMRI data.

“Harmonization” is a way to mitigate the measurement differences attributed to the scanner-, protocol-, or other site-related differences¹⁰⁻¹³. Harmonization of multi-site dMRI datasets can dramatically increase the statistical power of neuroimaging studies and enable comparative studies pertaining to several brain disorders^{8,14}. When harmonizing dMRI data from different sites or scanners, one is ultimately interested in preserving the variability purely associated with the biology or disease, and, at the same time, removing other sources of variability (i.e., intrinsic or acquisition-related variability of the scanners), which can obscure the effect of interest.

Apart from the above-mentioned variabilities, the problem of harmonizing datasets from multiple sites becomes even more challenging when one considers the variability that can arise at each step of the processing pipeline. Typical dMRI pre-processing pipeline involves multiple steps, such as denoising, correction for subject motion, and geometrical distortion correction due to eddy currents and susceptibility artifacts. However, no consensus thus far exists on which

processing steps to include, in which order, and which software to use for each step. Thus, minor changes in the pre-processing pipeline can bias the harmonization algorithm.

Various data harmonization methods have been proposed in the literature. These methods can be grouped into two main categories depending on the stage at which the harmonization is performed:

(1) *Harmonization of the dMRI data at the signal level*: These approaches use the dMRI data directly and apply the harmonization at the signal level early in the pre-processing pipeline, avoiding (or protecting against) additional bias downstream in data processing as well as subsequent data modeling. Any model fitting, tractography, or connectivity analysis can be applied in later stages of the processing pipeline after harmonization^{12,15–17}. On the other hand, the approaches in this category require groups of healthy subjects from multiple sites for learning scanner differences.

(2) *Statistical harmonization of diffusion maps*: These methods have been used in multiple fields, from genetics to functional MRI and recently in dMRI^{5,18–22}. They use the desired diffusion measures from multiple sites/scanners as input and apply statistical data pooling at the final stage of the processing pipeline to reduce the unwanted inter-site effects^{5,21,23,24}. The state of the art approach in this category is ComBat, a batch-effect correction tool used in genomics²², and adapted to diffusion MRI harmonization²¹. ComBat is a powerful and fast statistical data pooling tool, which can estimate an additive and a multiplicative site-effect parameter at each voxel, thus accounting for voxel-wise scanner differences. Combat harmonization can be applied per voxel but also over entire regions of interest.

While many studies emphasized the benefits of the harmonization at the signal level, there are multiple advantages of using statistical harmonization approaches especially when a sample size that guarantees a statistically representative sample from each scanner is included in the studies. Furthermore, statistical harmonization (e.g., ComBat) is fast and easy to employ, and a more practical option when only diffusion measures of individual subjects from different sites are available, but not necessarily the dMRI data. For example, multi-center studies and consortia might only be able to share the diffusion maps (i.e. fractional anisotropy (FA) maps) of individual subjects; which makes statistical harmonization the only way to remove site effects. Thus, statistical harmonization approaches are more commonly used, but are more vulnerable to the pre-processing effects. It is, therefore, very important to explore these approaches thoroughly. In this study, we explore the commonly used approach, ComBat, on its ability to harmonize multi-site diffusion maps (measures). It is a well-known fact that a good harmonization algorithm should preserve the inter-subject biological variability from each site while removing only the scanner-specific effects and be minimally sensitive to changes in the pre-processing steps. To test whether ComBat has these characteristics, we used dMRI data from five sites, including 542 individuals. We investigated two important aspects of using ComBat harmonization, which have not been well-examined in prior studies. 1) How well does ComBat preserve effect sizes when used for harmonization, and 2) What are the effects of using different software packages to estimate the diffusion tensor on the harmonization performance of ComBat. Consequently, this study aims to answer these questions by harmonizing FA across five sites using ComBat.

Methods:

Participants and diffusion MRI data acquisition.

This study utilized dMRI data collected from five sites. Overall, these included a group of schizophrenia patients comprising 291 individuals (mean [SD] age, 28.57 [13.02] years; 198 [68%] male), as well as a healthy control group comprising 251 individuals (mean [SD] age, 26.77 [12.09] years; 157 [63%] male). Refer to Table 1 for demographics and acquisition parameters of each site. IRB/Sample previously reported in Cetin-Karayumak et al., 2019⁸.

--- Insert Table 1 ---

Diffusion MRI data pre-processing.

A standardized processing procedure (eddy current and motion correction) was applied to all subjects using FSL's eddy correction software²⁵. Next, for each dMRI data in 5 sites, we fit the diffusion tensor with ordinary least squares fitting and computed FA using three software packages (FSL v. 6.0.1²⁶, MRtrix v. 3.0²⁷, and Slicer v. 4.10.2²⁸). To compare the whole brain white matter and major bundles across sites, the Illinois Institute of Technology (IIT) v. 4.1 Human Brain FA atlas in MNI space²⁹ was registered to the FA maps in subject space using *antsRegistrationSyn.sh*³⁰. Next, we warped the IIT white matter probabilistic atlas consisting of 17 bundles to the subject space. The probabilistic bundles were thresholded at 0.25, resulting in a total of 17 regions of interest (ROIs) of white matter fiber bundles. For each ROI, mean FA was computed by averaging FA across all voxels traversing the individual ROI. Additionally, average FA in the whole brain white matter was computed to complement the analysis. Mean FA of 17 ROIs + whole brain was used as input to ComBat to perform harmonization across the five sites.

--- Insert Figure 1 ---

ComBat Harmonization.

The ComBat model was introduced in the context of gene expression analysis by Johnson et al.²² and reformulated by Fortin et al.²¹ in the context of dMRI. Briefly, ComBat uses a Bayesian framework to estimate the additive and multiplicative site-effect hyperparameters at each voxel or ROI empirically. In this work, we utilized the Matlab version of ComBat software (<https://github.com/Jfortin1/ComBatHarmonization>) for harmonization of 17 ROIs + whole brain across five sites. Age, sex and disease status were added as biological covariates to the design matrix.

Post-analysis.

The ability of ComBat to perform accurate harmonization was tested by comparing the effect sizes between groups before and after harmonization. Effect sizes for FA differences in Cohen's d were computed between healthy females and males, as well as separately between all patients and controls to assess group differences. The mathematical definition of effect sizes in Cohen's d can be found in^{12,31,32}. According to Cohen's d criteria, small effect size is 0.2, medium is 0.5 and large is 0.8. The effect sizes were computed before and after ComBat harmonization at each site for comparative analysis.

Results:

Inter-group/biological variability was measured by the effect sizes of the FA differences between two groups within each site:

- (group i) healthy female and male controls;
- (group ii) all healthy controls and schizophrenia patients.

We conducted the following two experiments to assess whether the inter-group variability was preserved after harmonization. We note that the effect sizes before and after harmonization are respectively referred to as “before” and “after” in figures.

(*Experiment 1*). In the first experiment, the entire dMRI data pre-processing was conducted in a consistent manner with the same software across five sites. We assessed the effect sizes of FA differences of females and males (group i) after harmonization, where FA was computed using FSL for all sites and mean FA of 17 ROIs + the whole brain was given as input to the ComBat harmonization algorithm. Figure 2 depicts the effect sizes of each ROI and whole brain for site 1 and site 5, both before harmonization and after harmonization with ComBat (refer to supplementary Figure 1 for the other sites). At site 1, the effect sizes were preserved before and after harmonization for nearly all ROIs, barring a reduction in effect size in the negative direction for left inferior fronto occipital fasciculus (before harmonization: -0.25, after harmonization: -0.05). On the other hand, for site 5, the effect sizes for almost all ROIs were altered after harmonization, i.e. the effect sizes were amplified across many ROIs after harmonization with ComBat (e.g. see right cingulum - cingulate gyrus portion, before harmonization: -0.2, after harmonization: -0.5). For five ROIs (right inferior fronto occipital fasciculus, right corticospinal tract, left superior longitudinal fasciculus, left corticospinal tract, left cingulum - hippocampal portion), the sign of the effect size was flipped, indicating that after harmonization the direction of the FA differences in group i were reversed.

--- Insert Figure 2 ---

(*Experiment 2*). In the second experiment, we used different software packages to compute FA for each site to explore how subtle changes in the processing pipeline affect harmonization. The mean FA of 17 ROIs + the whole brain computed using different software packages for each

site was provided as input to the ComBat harmonization algorithm. Figure 3 depicts the effect sizes between all healthy controls and schizophrenia patients (group ii) of site 3, both before harmonization and after harmonization with ComBat (refer to Supplementary Figure 2 for the other sites). We observed that when different software was used for each site to compute FA, effect sizes were not preserved after harmonization and the absolute differences in effect sizes between before and after harmonization were as large as 0.7. In addition, the effect sizes of seven ROIs changed sign (see right superior longitudinal fasciculus, right inferior longitudinal fasciculus, right cingulum hippocampal, right cingulum - cingulate gyrus, right uncinate, left cingulum - hippocampal, fornix). For example, for the right cingulum - hippocampal portion, while the effect size before harmonization was 0.3, it became -0.2 after harmonization.

--- Insert Figure 3 ---

Discussion:

Harmonization of dMRI data along with a consistent processing pipeline is critical for reproducible neuroscientific research and provides increased statistical power. Statistical harmonization methods have been widely preferred in the literature for joint analysis of dMRI derived scalar maps due to their simplicity in neuroimaging applications. These harmonization methods such as Meta-analysis⁵ and ComBat²¹ statistically pool dMRI measures (FA, mean diffusivity, etc.) at the final stages of the processing pipeline. In this study, using a large dataset of 542 subjects from 5 sites, we evaluated ComBat in its ability to harmonize multi-site dMRI data by comparing the effect sizes (Cohen's d) in two groups, before and after harmonization. To this end, we conducted two experiments. In both experiments, we applied the same pre-processing pipeline for all sites, except that in experiment 2, we used different software packages to fit the diffusion tensor and compute FA for each site.

In experiment 1, we observed that even when all processing steps, as well as diffusion tensor estimation software were identical for all sites, ComBat harmonization still seemed to alter the inter-group differences and, in some cases, flipped the direction of effect sizes. This indicates that after harmonization, the direction of the differences was reversed. This might be due to ComBat's optimization procedure that assumes that the site-effect parameters follow a particular parametric prior distribution (Gaussian and Inverse-gamma), which might not be true in all scenarios. We note that our findings support results in structural neuroimaging studies demonstrating that the standard ComBat approach might not necessarily preserve biological variability^{33–35}. Additionally, we notice that at site 5, the data was acquired on a 1.5T scanner, whereas at site 1 it was acquired on a 3T scanner. Thus, the scanner differences are potentially highly non-linear which are not well captured by ComBat.

In experiment 2, we investigated the effects of additional sources of variability that can possibly arise at any step of the dMRI processing pipeline, when different software packages or versions are used for data pooling using ComBat (for example, when already processed data is shared for joint analysis across consortia or sites). We first applied FSL's eddy to all sites for motion and eddy correction. We later used three different software packages (FSL, MRtrix, Slicer) to estimate the diffusion tensor and compute FA at different sites. We consistently used ordinary least squares fitting for diffusion tensor computation in all three software packages. These software packages perform pre-processing in slightly different ways and order (e.g., signal normalization by b0, denoising, thresholding after normalization to restrict data between 0 and 1, fixing negative eigenvalues, etc.). In this experiment, we observed that group differences were not preserved after harmonization. Thus, minor differences in pre-processing produced medium to large differences in the data after harmonization. This might be due to the fact that, the data at site 3 was interpolated on the scanner (in-plane resolution was 0.94 x 0.94 mm²), potentially causing the data to have many more negative values or negative eigenvalues and

more partial volume effects due to smoothing. Consequently, minor differences in the pre-processing steps resulted in non-linear changes in the input data, thereby causing ComBat to perform in an unexpected manner.

We note that, we only evaluated the standard ComBat method²¹ that is available as open-source (<https://github.com/Jfortin1/ComBatHarmonization>). The recent versions of ComBat, such as the generalized additive model/GAM - ComBat³⁴ might mitigate some of the problems; however, it remains to be tested on large dMRI datasets like the one used in this work.

Conclusion:

This study thoroughly explored the powerful ComBat harmonization in a large dMRI data of 542 individuals. We demonstrated that while ComBat harmonization preserves the effect sizes of between-group FA differences in some sites, it alters the between-group differences (significantly in some cases) and in some ROIs flips the sign of the effect sizes. We thus suggest caution while using harmonization methods like ComBat. Most importantly, we recommend using a consistent pre-processing pipeline in multi-site dMRI studies, more specifically, not to combine dMRI measures processed with different software packages.

Acknowledgments:

The authors have been supported by the following grants: NIH R01 MH119222 (Rathi), R01 MH102377 and K24 MH110807 (Kubicki), Swiss National Science Foundation grant 152619 (Walther). The project also acknowledges that the research has partly been supported by the BWH Program for Interdisciplinary Neuroscience, through a gift from Lawrence and Tiina Rand (Cetin-Karayumak).

Table 1 – All dMRI data were collected as part of separate, individual studies. Basic demographics and dMRI acquisition information of each site are provided in the table.

Name of the study/location	Sample #	Age (mean+/- std)	Gender	dMRI data
Site 1 Siemens Magnetom Trio 3T	31H 46S	40.74+/-11.31 38.37+/-11.43	22M, 9F 28M, 18F	b=1300 s/mm ² 42 directions resolution=2x2x2 mm ³
Site 2 GE Signa HDx	70H 68S	30.3+/-13.44 35.07+/-14.16	48M, 22F 51M, 17F	b=900 s/mm ² 59 directions resolution=1.7x1.7x1.7 mm ³
Site 3 GE HDx 3T	69H 63S	21.50+/-5.048 21.31+/-4.94	45M, 24F 46M, 17F	b=1000 s/mm ² 31 directions resolution=0.94x0.94x2.5 mm ³
Site 4 Siemens Tim Trio 3T	25H 42S	37.04+/-11.14 37.67+/-10.83	16M, 9F 32M, 10F	b=1000 s/mm ² 60 directions resolution = 2x2x2 mm ³
Site 5 Siemens Sonata 1.5T	56H 72S	16.5+/-1.74 17.17+/-1.66	26M, 30F 41M, 31F	b=1000 s/mm ² 60 directions resolution=2.5x2.5x2.5 mm ³
Total	251H 291S	26.77+/-12.09 28.57+/-13.02	157M, 94F 198M, 93F	
Abbreviations: H: Healthy Controls, S: Schizophrenia Patients; F: females; M: males.				

Figure 1 – Pre-processing steps of dMRI data collected from 5 sites: (0) All dMRI data is axis aligned, centered and eddy current corrected using the same pipeline; (1) for all subjects, FA is computed using three software with ordinary least squares fitting; (2) FA atlas is registered to each subject space using antsRegistration and probabilistic maps for 17 major bundles and whole brain white matter tissue map is warped to the subject space; (3) mean FA of whole brain white matter and 17 bundles are computed and given as input to ComBat harmonization.

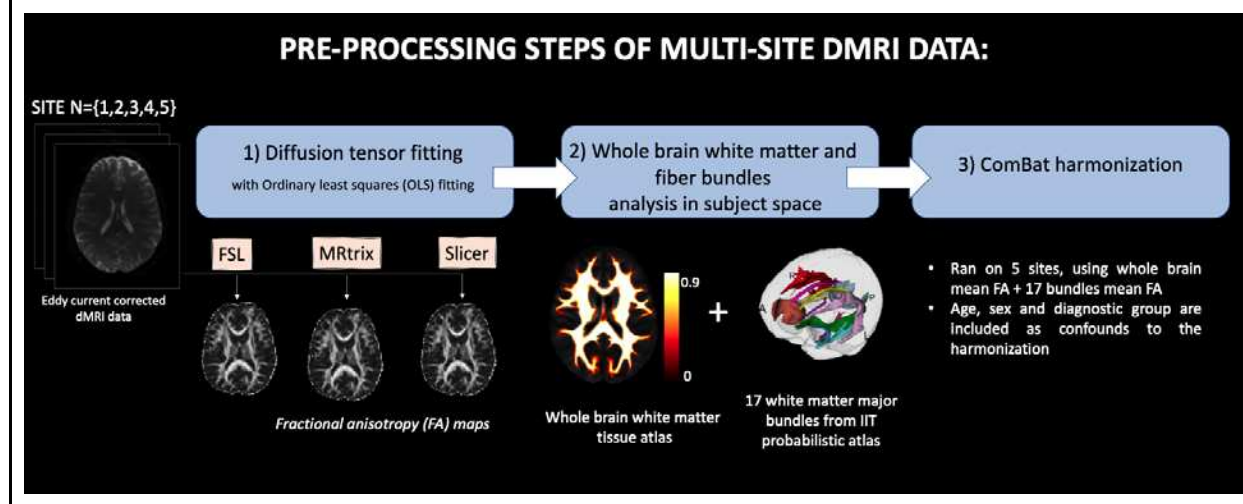


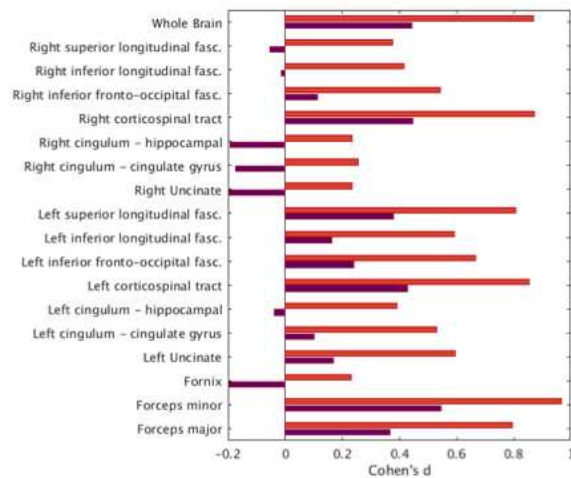
Figure 2 – Effect sizes (Cohen's d) of FA differences between female and male healthy controls are computed to assess whether the inter-subject variability is preserved after harmonization. The results are shown for site 1 and site 5 for the whole brain and 17 bundles. Purple bar represents effect sizes between two groups before harmonization (“before”). Orange bar represents effect sizes after harmonization, where FSL's FA was given as input to the harmonization (“after”). Data pre-processing was done in a consistent manner with the same software. At site 1, the effect sizes were almost preserved before and after harmonization (barring a reduction in effect size for left-inferior-fronto-occipital-fasciculus). However, at site 5, we observe that the effect sizes for almost all fiber bundles are altered after harmonization. Notably, ComBat seems to have increased the effect sizes across many fiber bundles. Most importantly, in a few fiber bundles, the sign of the effect size is flipped, indicating that after harmonization the direction of the differences was reversed, which can dramatically reduce reliability of the results.



Figure 3 – Effect sizes (Cohen’s d) of FA differences between healthy controls and schizophrenia patients are computed to assess whether the inter-subject variability is preserved after harmonization. The results are shown for Site 3 for the whole brain and 17 bundles. Purple bar represents effect sizes between two groups before harmonization (“before”). Orange bar represents effect sizes after harmonization, where FSL’s FA was given as input to the harmonization (“after”). To explore how subtle change in the pre-processing pipeline affects harmonization, different software to compute FA were given as input to the ComBat harmonization algorithm. After harmonization, effect sizes were not preserved such that absolute differences between “before” and “after” effect sizes were up to 0.7. In addition, we point out to the bundles where the effect sizes changed direction (e.g. right-cingulum-hippocampal). These results demonstrate that inter-subject variability was not preserved after harmonization and subtle changes in the pre-processing pipeline induced much larger biases, which were not properly accounted for by ComBat.

Effect sizes between healthy controls and schizophrenia patients

SITE 3



Before: Effect sizes between controls and patients before harmonization

After: Effect sizes between controls and patients after harmonization

Note: different software was selected to compute FA in each site

References:

1. Kubicki, M. *et al.* A review of diffusion tensor imaging studies in schizophrenia. *J. Psychiatr. Res.* **41**, 15–30 (2007).
2. Shenton, M. E. *et al.* A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain Imaging and Behavior* vol. 6 137–192 (2012).
3. Melonakos, E. D. *et al.* Voxel-based morphometry (VBM) studies in schizophrenia-can white matter changes be reliably detected with VBM? *Psychiatry Res.* **193**, 65–70 (2011).
4. Poldrack, R. A. *et al.* Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* **18**, 115–126 (2017).
5. Thompson, P. M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
6. Addington, J. *et al.* North American Prodrome Longitudinal Study (NAPLS 3): Methods and baseline description. *Schizophr. Res.* (2020) doi:10.1016/j.schres.2020.04.010.
7. Casey, B. J. *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
8. Cetin-Karayumak, S. *et al.* White matter abnormalities across the lifespan of schizophrenia: a harmonized multi-site diffusion MRI study. *Mol. Psychiatry* (2019) doi:10.1038/s41380-019-0509-y.
9. Seitz, J. *et al.* Investigating Sexual Dimorphism of Human White Matter in a Harmonized, Multisite Diffusion Magnetic Resonance Imaging Study. *Cerebral Cortex* (2020) doi:10.1093/cercor/bhaa220.
10. Mirzaalian, H. *et al.* Inter-site and inter-scanner diffusion MRI data harmonization. *Neuroimage* **135**, 311–323 (2016).
11. Mirzaalian, H. *et al.* Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging Behav.* **12**, 284–295 (2018).
12. Cetin Karayumak, S. *et al.* Retrospective harmonization of multi-site diffusion MRI data

- acquired with different acquisition parameters. *Neuroimage* **184**, 180–200 (2019).
13. Pinto, M. S. *et al.* Harmonization of Brain Diffusion MRI: Concepts and Methods. *Front. Neurosci.* **14**, (2020).
14. Kelly, S. *et al.* Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Mol. Psychiatry* **23**, 1261–1269 (2018).
15. Karayumak, S. C., Kubicki, M. & Rathi, Y. Harmonizing Diffusion MRI Data Across Magnetic Field Strengths. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* 116–124 (2018) doi:10.1007/978-3-030-00931-1_14.
16. Tax, C. M. *et al.* Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms. *Neuroimage* **195**, 285–299 (2019).
17. Ning, L. *et al.* Cross-scanner and cross-protocol multi-shell diffusion MRI data harmonization: Algorithms and results. *Neuroimage* **221**, 117128 (2020).
18. Kochunov, P. *et al.* 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 (2014).
19. Thompson, P. M. *et al.* ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10**, 100 (2020).
20. Zhang, G. *et al.* ENIGMA-Viewer: interactive visualization strategies for conveying effect sizes in meta-analysis. *BMC Bioinformatics* **18**, 253 (2017).
21. Fortin, J.-P. *et al.* Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* **161**, 149–170 (2017).
22. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
23. Pohl, K. M. *et al.* Harmonizing DTI measurements across scanners to examine the

development of white matter microstructure in 803 adolescents of the NCANDA study.

Neuroimage **130**, 194–213 (2016).

24. Timmermans, C. *et al.* Potential of a statistical approach for the standardization of multicenter diffusion tensor data: A phantom study. *J. Magn. Reson. Imaging* **49**, 955–965 (2019).
25. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
26. Behrens, T. E. J. *et al.* Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* **50**, 1077–1088 (2003).
27. Veraart, J., Sijbers, J., Sunaert, S., Leemans, A. & Jeurissen, B. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. *Neuroimage* **81**, 335–346 (2013).
28. Norton, I. *et al.* SlicerDMRI: Open Source Diffusion MRI Software for Brain Cancer Research. *Cancer Res.* **77**, e101–e103 (2017).
29. Varentsova, A., Zhang, S. & Arfanakis, K. Development of a high angular resolution diffusion imaging human brain template. *Neuroimage* **91**, 177–186 (2014).
30. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
31. Statistical Power Analysis for the Behavioral Sciences. (1977) doi:10.1016/c2013-0-10517-x.
32. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. (2013) doi:10.4324/9780203771587.
33. Garcia-Dias, R. *et al.* Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners. *Neuroimage* **220**, 117127 (2020).
34. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging

patterns throughout the lifespan. *Neuroimage* **208**, 116450 (2020).

35. Da-ano, R. *et al.* Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci. Rep.* **10**, 1–12 (2020).