

Harmonizing multisite neonatal diffusion-weighted brain MRI data for developmental neuroscience

Alexandra F. Bonthron¹, Manuel Blesa Cábez^{2,3}, A. David Edwards¹, Jo V. Hajnal¹, Serena J. Counsell¹, James P. Boardman^{2,3}

1. Centre for the Developing Brain, School of Biomedical Engineering and Imaging Sciences, King's College London

2. Centre for Reproductive Health, Institute for Regeneration and Repair, University of Edinburgh

3. Centre for Clinical Brain Sciences, University of Edinburgh

Corresponding Author: Dr Alexandra F. Bonthron; email: alexandra.bonthron@kcl.ac.uk

Abstract

Large diffusion-weighted brain MRI (dMRI) studies in neonates are crucial for developmental neuroscience. Our aim was to investigate the utility of ComBat, and empirical Bayes tool for multisite harmonization, for removing site effects from white matter (WM) dMRI measures in healthy infants born 37-42+6 weeks from the Theirworld Edinburgh Birth Cohort (n=86) and Developing Human Connectome Project (n=287).

Skeletonized fractional anisotropy (FA), mean, axial and radial diffusivity (MD, AD, RD) maps were harmonized. The differences between voxel-wise metrics, skeleton means and histogram widths (5th-95th percentile) were assessed before and after harmonization, as well as variance associated with gestational age at birth.

Before harmonization, large cohort differences were observed in all measures. Harmonization removed all voxel-wise differences from MD maps and all metric means and histogram widths, however small voxel-wise differences (<1.5% of voxels) remained in FA, AD and RD. We detected significant relationships between GA at birth and all metrics. When comparing single site and multi-site harmonized datasets of equal sample sizes, harmonized data resulted in smaller standardized regression coefficients.

ComBat will enable unprecedented sample sizes in developmental neuroscience, offering new horizons for biomarker discovery and validation, understanding typical and atypical brain development, and assessment of neuroprotective therapies.

1. Introduction

Magnetic resonance imaging (MRI) studies in neonates are crucial for assessing brain development and elucidating mechanisms underlying typical and atypical neurodevelopmental

outcomes. However, neurodevelopment is complex and influenced by multifactorial, often interacting neurobiological, (epi)genetic and environmental processes (Sonuga-Barke, 2023). Large, representative samples are required to disentangle these processes; this can only be achieved by using data acquired from multiple sites. Several large neonatal brain MRI cohorts exist that are now available to the research community (Boardman et al., 2020; Edwards et al., 2022; Howell et al., 2019; Soh et al., 2014); these provide new opportunities to combine datasets acquired across multiple sites. This would create unprecedented sample sizes to characterize typical population variation, delineate neurodevelopmental risk and resilience factors, identify robust and generalizable predictors of outcomes, investigate effects of (epi)genetic variation and test novel neuroprotective therapies in at-risk groups. However, there is an urgent need to identify robust strategies for combining data acquired across multiple locations.

Diffusion-weighted MRI (dMRI) measures the displacement of water molecules in tissue over time and is used to characterize microstructure in the developing brain (Pecheva et al., 2018). Mathematical models of the dMRI signal provide quantitative information which can be used to infer characteristics of the underlying tissue. The most used model is the diffusion tensor model (DTI), which provides rotationally invariant scalar metrics (fractional anisotropy [FA], and mean, axial and radial diffusivity [MD, AD, RD]) that characterize the diffusion properties of a tensor fitted to each voxel (Basser et al., 1994). DTI metrics can be calculated from standard clinical diffusion MRI sequences (Le Bihan et al., 2001) which have been widely used in neonatal dMRI research. Tensor-derived measures of white matter microstructure may be altered in neonates at risk of altered brain development such as those born preterm (Dibble et al., 2021; Vaher et al., 2022), with hypoxic ischemic encephalopathy (Dibble et al., 2020; Ward et al., 2006) and with congenital heart disease (Karmacharya et al., 2018; Miller et al., 2007; Mulkey et

al., 2014). These measures are associated with both clinical and environmental risk factors such as nutrition, infection, and the perinatal stress environment (Barnett et al., 2018; Blesa et al., 2019; Boardman & Counsell, 2020; Demers et al., 2021; Lautarescu et al., 2019; Lean et al., 2022; Stoye et al., 2020; Sullivan et al., 2020; Wheeler et al., 2022) and childhood neurodevelopmental outcomes in both typical and atypical populations (Dimitrova et al., 2020; K. Feng et al., 2019; Lautarescu et al., 2022; Salvan et al., 2017; Tusor et al., 2012; Ullman et al., 2015). In addition, dMRI measures have been used as markers of treatment efficacy in trials of neuroprotective therapies for neonatal populations (Azzopardi et al., 2016; Law et al., 2021; Moltu et al., 2024; O’Gorman et al., 2015; Poppe et al., 2022; Porter et al., 2010). However, dMRI-derived measures vary by both scanner and acquisition, meaning it is challenging to combine data from multiple centres.

Automated quantitative analysis of dMRI measures across the brain often requires registration to a common space, which is challenging given the large variation in brain size, shape and signal intensity across the neonatal period (Dubois et al., 2021). Methods such as tract-based spatial statistics (TBSS) (Smith et al., 2006) including its optimization for neonatal datasets (Ball et al., 2010), aim to overcome some of these challenges by non-linearly registering data, isolating a white matter ‘skeleton’ comprising of voxels with the highest FA, and projecting data from individuals onto this skeleton to ensure spatial alignment and remove contamination from non-white matter structures. Tensor-based registration algorithms such as DTI-ToolKit (DTI-TK) (Zhang et al., 2006) provide superior registration of neonatal dMRI for TBSS (Wang et al., 2011). Skeletonised DTI metrics in neonates have been analysed on a voxel-wise basis using permutation testing to identify spatially homogenous regions of alterations across the white matter skeleton (Barnett et al., 2018; Blesa et al., 2019; Mulkey et al., 2014; Porter et al., 2010).

Mean skeletonised DTI metrics can also be extracted from the whole skeleton or from a-priori hypothesised regions of the skeleton identified from a white matter atlas. However, both voxel-wise and analyses of metric averages assume homogeneity in the location and the extent of white matter alterations. Peak-width of skeletonized water diffusion metrics is a method originally developed to assess small vessel disease in adults (Baykara et al., 2016). This method measures the width of the histogram of values within the white matter skeleton (5th to 95th percentile). Blesa and colleagues (Blesa et al., 2020) identified correlations between MD, AD and RD histogram widths and gestational age (GA) at birth; furthermore, peak-width of skeletonised dMRI measures accurately classifies images according to GA at birth. These metrics are of particular interest for large scale studies of the neonatal brain because they characterize generalized white matter maturation and will capture changes that may be heterogenous in location and extent. In adults with small vessel disease, histogram widths were highly reproducible across MRI scanners and field strengths (Baykara et al., 2016); however, it is not known if these are reproducible across MR scanners in the newborn brain, or whether harmonization is needed to make valid inter-site comparisons.

ComBat (Johnson et al., 2007) is a data harmonization tool originally designed to remove batch effects from genomic data and is an effective harmonization tool for adult DTI data (Fortin et al., 2017). However, the effect of ComBat data harmonization on voxel-wise, mean and histogram widths across the white matter skeleton in healthy typically developing infants has not been comprehensively assessed.

The aims of this study were to assess (i) whether FA, MD, AD and RD histogram widths are comparable across scanning sites and (ii) the utility ComBat for removing site effects from DTI measures of white matter development in healthy typically developing infants born at 37-42

weeks gestational age ('term'). We assessed differences between sites on a voxel-wise basis and across the mean and histogram width before and after harmonization. To investigate the effect of harmonization on correlation effects, we assessed the relationship between dMRI metrics and GA at birth before and after harmonization. GA at birth was chosen as birth earlier in the term period (e.g. 37-38 weeks) is associated with altered neonatal white matter development (Broekman et al., 2014a; Gale & Grant et al., 2022; Jin et al., 2019; Ou et al., 2017a) lower neurodevelopmental scores (Hua et al., 2019; Rose et al., 2013).

2. Methods

The National Research Ethics Service Research Ethics Committees in West London (Developing Human Connectome Project (dHCP) 14/LO/1169) and South East Scotland (Theirworld Edinburgh Birth Cohort (TEBC) 16/SS/0154) provided ethical approvals.

In accordance with the declaration of Helsinki, informed written parental consent was obtained before MRI.

2.1 Participants

Preprocessed dMRI data from a subset of typically developing healthy infants at low risk of altered brain development born ≥ 37.0 weeks from Theirworld Edinburgh Birth Cohort (Boardman et al., 2020) and the Developing Human Connectome Project (Edwards et al., 2022) were used.

2.2 MRI acquisition

Infants from both cohorts underwent brain MRI during natural sleep with monitoring of pulse oximetry, electrocardiography, and temperature. All scans were supervised by a doctor or nurse trained in neonatal resuscitation and MR procedures.

Images acquired at each centre were reviewed by neuroradiologists experienced in neonatal brain imaging and all infants included in this analysis had no evidence of major incidental findings.

2.2.1 Developing Human Connectome Project

Brain MRI was performed on a Philips Achieva 3 Tesla system (Best, Netherlands) situated in the neonatal intensive care unit at St. Thomas' Hospital using a 32-channel neonatal head coil and neonatal positioning device (Hughes et al., 2017). Acoustic protection consisted of earplugs made from silicone-based putty placed in the external auditory meatus (President Putty, Coltene Whaledent, Mahwah, NJ), neonatal earmuffs (MiniMuffs; Natus Medical, Middleton, WI) and an acoustic hood placed over the infant.

dMRI was acquired with a high angular resolution diffusion multi-shell protocol designed for the neonatal brain (TR/TE 3800/90 ms, multiband acceleration factor 4, sensitivity encoding in-plane acceleration factor 1.2, in-plane resolution 1.5×1.5 mm, slice thickness 3 mm, 1.5 mm overlap, 300 volumes, diffusion gradient encoding: $b = 0$ s/mm² (n = 20), $b = 400$ s/mm² (n = 64), $b = 1000$ s/mm² (n = 88), $b = 2600$ s/mm² (n = 128) with 4x interleaved phase encoding)(Hutter et al., 2018).

2.2.2 Theirworld Edinburgh Birth Cohort

Brain MRI was performed on a Siemens MAGNETOM Prisma 3 Tesla system (Erlangen, Germany) using a 16-channel pediatric head and neck coil. Acoustic protection consisted of flexible ear plugs and neonatal earmuffs (MiniMuffs; Natus Medical, Middleton, WI).

High angular resolution dMRI was acquired in 2 separate acquisitions to reduce the time needed to re-acquire any data lost to motion artifacts. The first acquisition consisted of 8 baseline volumes ($b = 0 \text{ s/mm}^2$ [b0]) and 64 volumes with $b = 750 \text{ s/mm}^2$, and the second multi-shell acquisition consisted of 8 b0, 3 volumes with $b = 200 \text{ s/mm}^2$, 6 volumes with $b = 500 \text{ s/mm}^2$, and 64 volumes with $b = 2,500 \text{ s/mm}^2$ (Acquisition parameters for both sequences: TR/TE 3500/78.0 ms; multiband acceleration factor 2, in-plane acceleration factor 2, in-plane resolution 2mm x 2mm, slice thickness 2mm).

2.3 dMRI preprocessing

dMRI from each cohort were preprocessed according to local procedures, described below.

2.3.1 Developing Human Connectome Project

dMRI underwent parallel imaging reconstruction, denoising (Cordero-Grande et al., 2019; Veraart et al., 2016), removal of Gibbs ringing artefacts (Kellner et al., 2016), and correction for motion and image distortion using Spherical Harmonics and Radial Decomposition (Christiaens et al., 2021). The $b=1000 \text{ s/mm}^2$ shell was extracted for further analysis.

2.3.2 Theirworld Edinburgh Birth Cohort

The dMRI acquisitions underwent denoising (Veraart et al., 2016) and correction for motion and image distortion using outlier replacement and slice-to-volume registration (Andersson et al., 2003, 2016, 2017; Andersson & Sotiropoulos, 2016; Smith et al., 2004). Bias field inhomogeneity correction was performed by calculating the bias field of the mean b0 volume and applying the correction to all the volumes (Tustison et al., 2010). The first acquisition consisting of 64 directions at $b=750 \text{ s/mm}^2$ was used for further analysis.

2.4 dMRI image processing

b=1000s/mm² data from the dHCP and b=750s/mm² were processed using the FMRIB software library (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>; FSL)(Smith et al., 2004) and DTI-TK (www.dti-tk.sourceforge.net)(Zhang et al., 2006). Diffusion tensors were calculated on a per voxel basis. Tensor images were registered to the Edinburgh Neonatal Atlas tensor template (Blesa et al., 2016, 2020) using DTI-TK. FA, MD, AD and RD maps were calculated.

A mean FA image was calculated and a mask of the mean FA skeleton was derived by perpendicular non-maximum suppression and an FA threshold of 0.15. The study-specific skeleton mask was multiplied with the custom Edinburgh neonatal atlas template skeleton mask to remove grey matter regions and fibres passing through the cerebellum, the brainstem and the subcortical grey matter (Blesa et al., 2020), based on the custom mask created by Baykara and colleagues (Baykara et al., 2016). FA, MD, AD, and RD maps for each dataset were projected onto this skeleton. For each infant, the skeleton mean and the width of the histogram (the difference between the 95th and 5th percentile voxel) for FA, MD, AD and RD were calculated.

2.5 Data harmonization

2.5.1 ComBat

ComBat (Johnson et al., 2007) implemented in NeuroHarmonize (Pomponio et al., 2020) was used to remove the effect of site of acquisition from the skeletonized FA, MD, AD and RD maps. ComBat is an empirical Bayes method which linearly models the additive and multiplicative effects of site on feature values to minimize the variance associated with site and preserve the variance most associated with variables of interest included in the ComBat model (Fortin et al.,

2017). In this analysis, the variance associated with GA at birth, GA at scan and sex were preserved.

2.6 Statistical analysis

All analyses except for the voxel-wise analyses were undertaken in R version 4.0.3. Differences in demographic variables between cohorts were assessed with Mann-Whitney U tests (GA at birth, GA at scan) and χ^2 (sex).

2.6.1 Voxel-wise analysis

To investigate the relationship between voxel-wise dMRI metrics, site and GA at birth before and after harmonization, voxel-wise permutation testing (10000 permutations) was performed using Randomise in FSL (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise>). Significant voxels are displayed on the mean FA image and T-statistic ranges and percentage of significant voxels are reported.

2.6.1.1 Differences between sites

General linear models (GLM) were used to assess the differences in dMRI metrics between dHCP and TEBC before and after harmonization with GA at birth, GA at scan and sex as covariates. Voxel-wise values before and after harmonization were extracted across the whole skeleton and mean difference plots created.

2.6.1.2 Associations with GA at birth

To assess the relationship between voxel-wise dMRI metrics and GA at birth before and after harmonization, a GLM was used with GA at scan and sex as covariates.

A post-hoc analysis was undertaken to investigate the relationship between postnatal age (calculated as: GA at scan – GA at birth) and FA in regions where lower GA at birth was associated with higher FA. Mean FA values for each infant were extracted from voxels where lower FA was associated with higher GA at birth in the harmonized dataset. The association between extracted FA values and postnatal age was assessed adjusting for GA at scan and sex.

To assess the impact of data harmonization on sensitivity, we conducted voxel-wise permutation tests with five subsamples of non-harmonized data (dHCP alone, n=86 in each) and five subsamples of the full cohort of harmonized TEBC and dHCP (n=86 each). Significant voxels were extracted and the overlap of significant voxels across subsamples of un-harmonized data and the combined harmonized data.

2.6.2 Mean and histogram width analyses

2.6.2.1 Differences between sites

Multiple linear regressions were used to characterize the effect of site on dMRI metric means and histogram widths before and after harmonization with GA at birth, GA at scan and sex included as covariates. Standardised regression coefficients (β), standard errors and p-values are reported. The GVLMA package (Peña & Slate, 2006) assessed the validity of linear regressions, and for models where assumptions were violated, analyses were rerun with robust regression using fast-s algorithms (Lourenço et al., 2011).

2.6.1.2 Associations with GA at birth

The relationship between GA at birth and dMRI metric means and histogram widths was assessed (i) before harmonization in each site separately, (ii) before harmonization without a site covariate, (iii) before harmonization with a covariate and (iv) after harmonization using the procedures described above with GA at scan and sex included as covariates.

To assess the impact of data harmonization on standardised regression coefficients, a permutation analysis was conducted with 10,000 iterative subsamples of the un-harmonized dHCP (n=86), and 10,000 iterative subsamples of the harmonized dHCP and TEBC data (n=86) with robust regressions to assess the relationship between GA at birth and dMRI metrics adjusting for GA at scan and sex. Mean standardised regression coefficients, 95% confidence intervals and standard deviations were calculated for each metric. Independent samples t-tests were used to compare standardised regression coefficients between the dHCP alone and harmonized dHCP and TEBC.

2.7 Data availability

Data from the dHCP is from the 3rd neonatal data release (<https://www.developingconnectome.org/>) (Edwards et al., 2022). Requests for TEBC anonymized data will be considered under the study's Data Access and Collaboration policy and governance process (<https://www.ed.ac.uk/centre-reproductivehealth/tebc/about-tebc/for-researchers/data-access-collaboration>). Scripts used to prepare data for harmonization are included as a supplement.

3. Results

3.1 Demographics

The cohorts did not differ significantly for male:female proportion or GA at birth. GA at scan was higher in TEBC compared to dHCP (Table 1).

3.2 Combat

3.2.1 Voxel-wise metrics

Voxel-wise differences before and after harmonization are summarised in Figure 1. Before harmonization, 44% of FA voxels within the white matter skeleton differed between sites, with higher FA values in central white matter regions in the dHCP, and higher FA values in peripheral white matter structures in TEBC (t-statistic range -14.32-19.01). 88% of MD (t-statistic range -11.51-27.89), 73% of AD (t-statistic range -9.95-30.90) and 86% of RD voxels (t-statistic range -12.67-23.84) within the white matter skeleton differed between sites, with higher values in TEBC compared to dHCP.

After harmonization with ComBat, 0.2% of FA voxels were different between sites (t-statistic range -4.13-5.28), with higher values in TEBC in the right cerebellar peduncle. 0.4% of AD voxels differed between sites (t-statistic range -5.73-5.90), with higher AD in the left occipital white matter in the dHCP. 1.2% of RD voxels differed between sites (t-statistic range -5.86-4.05), with higher values in bilateral cerebellar peduncles in the dHCP. There were no significant differences in MD between sites (t-statistic range -5.82-4.94).

3.2.2 Mean and histogram widths

dMRI metric means and histogram widths before and after harmonization are summarised in Table 2 and supplementary figures 1 and 2. Before harmonization, there was a significant effect

of site on mean MD, AD and RD and all dMRI metric histogram widths in the white matter skeleton. After harmonization, there was no effect of site on dMRI metric means or histogram widths (Table 3).

3.2.3 Relationship with gestational age at birth

3.2.3.1 Voxel-wise metrics

Voxel-wise correlations with GA at birth before and after harmonization are summarised in Figure 2. Before harmonization, 28% of FA voxels within the white matter skeleton were associated with GA at birth. Older GA at birth was associated with higher FA values associated in the centrum semiovale, thalamus and corpus callosum and lower FA in the left posterior and retrolenticular parts of the internal capsule and bilateral cerebral peduncles (t-statistic range -8.37-6.59). 86% of MD (t-statistic range -9.89-4.45), 88% of AD (t-statistic range -10.18-3.96) and 80% of RD voxels (t-statistic range -13.00-4.55) were negatively associated with GA at birth.

After harmonization with ComBat, 22% of FA voxels were associated with GA at birth. Older GA at birth was associated with higher FA values associated in the centrum semiovale, and lower FA in the left retrolenticular part of the internal capsule (t-statistic range -6.70-6.71). Post-hoc analyses revealed FA values in the cluster within the left retrolenticular part of the internal capsule were significantly positively associated with postnatal age (β [SE]=0.358 [0.053], $p < 0.001$), adjusting for GA at scan and sex. 46% of MD voxels (t-statistic range -10.33-4.59), 50% of AD voxels (t-statistic range -8.42-5.09), 39% of RD voxels (t-statistic range -10.48-4.92) and were negatively associated with GA at birth across the supratentorial white matter.

When iteratively assessing voxel-wise correlations with GA at birth, t-statistic ranges were similar for TEBC (n=86), five subsets of the dHCP (n=86), and five subsets of harmonized TEBC and dHCP (n=86) (Figure 3). Significant voxels identified in each iteration of dHCP and harmonized TEBC and dHCP are shown in Figure 4.

3.2.3.2 Mean and histogram width

Table 4 summarises the relationship between GA at birth and dMRI metric means and histogram widths in: TEBC, dHCP, the combined dataset with no correction for site, the combined dataset with site as a covariate (linear correction) and the combined dataset after ComBat harmonization. Standardised regression coefficients were similar across TEBC, dHCP and the harmonized dataset, however statistical power is higher in the combined dataset, revealing significant associations between GA at birth and mean MD, AD and RD as well as histogram widths across all metrics. Figure 5 shows the relationship between GA at birth and mean FA and histogram width in TEBC, dHCP, both datasets after ComBat harmonization and both datasets with site as a linear covariate in the regression.

In permutation tests, standardised regression coefficients for gestational age at birth were significantly lower in the combined harmonized data for all metrics except FA histogram width (Table 5).

4. Discussion

In this study we assessed the utility of ComBat for data harmonization across skeletonized DTI metrics in the neonatal brain. Before data harmonization, large differences in voxel-wise, mean and histogram widths were observed between two cohorts of healthy control infants who underwent neonatal brain MRI using different scanning hardware and sequence parameters. Harmonization removed all voxel-wise differences from mean diffusivity maps, however small differences (<1.5% of voxels in the white matter skeleton) remained in FA, AD and RD. Harmonization also removed significant differences in skeletonised DTI metric means and histogram widths. In the large combined harmonized dataset, we were able to detect significant relationships between GA at birth and mean and histogram widths of DTI metrics even though the study population comprised term infants only. However, when comparing single site and multi-site harmonized datasets of equal sample sizes, harmonized data resulted in smaller standardised correlation coefficients for GA at birth.

Large differences in voxel-wise DTI metrics as well as skeletonised mean (except FA) and histogram widths were observed between sites, however these were removed with ComBat harmonization. Voxel-wise mean-difference plots, and skeleton means and histogram widths, of DTI metrics after harmonization remained different between cohorts, which reflects the significantly higher GA at scan in TEBC. When adjusting for GA at scan, GA at birth and sex, there were no significant effect of site on DTI metric means or histogram widths. When combining data from multiple sites, standardizing preprocessing and quality control strategies can

reduce variation between sites, however differences related to data acquisition are likely to impact extracted metrics (Galdi et al., 2024). ComBat has been used to harmonize DTI metric maps from very preterm infants imaged with different sequences on the same MRI system (Parikh et al., 2021), region of interest measures in term and preterm infants (Galdi et al., 2018), and DTI metrics along white matter tractography reconstruction in infants with congenital heart disease (Meyers et al., 2022). We add to this literature by demonstrating that ComBat successfully harmonizes voxel-wise DTI metrics as well as mean values and histogram widths from the white matter skeleton. This provides novel opportunities to retrospectively combine neonatal dMRI cohorts and overcome sample size limitations often inherent to neonatal neuroimaging research (Korom et al., 2022) and undertake large meta-analyses of neonatal neuroimaging studies. In addition, these tools may be used in clinical trials of neuroprotective agents in neonates which use dMRI as an outcome, but often recruit across multiple research centers to reach adequate sample sizes.

After harmonization with ComBat and adjustment for GA at scan, GA at birth and sex, small differences in voxel-wise FA, AD, RD were found between sites. These differences may represent subtle differences between cohorts not captured in the covariates included in this analysis. Fortin and colleagues reported no white matter voxels were associated with site after ComBat harmonization of FA and MD maps in adults (Fortin et al., 2007). However, the authors used a Bonferroni correction without threshold-free cluster enhancement to control the family-wise error rate, which is more conservative than the correction used in our study and may account for the differing results. Further work with larger multi-site samples is required to investigate differences between neonatal cohorts after data harmonization.

When iteratively examining the relationship between GA at birth and mean and histogram widths of DTI metrics across 10,000 subsets of 86 babies from dHCP data and harmonized dHCP and TEBC, regression coefficients for all metrics (except FA histogram width) were significantly higher in the dHCP alone subsets. Harmonization with ComBat may therefore slightly reduce statistical power when compared to a dataset of equivalent size from one site. However, it is important to note that across the whole harmonized sample, regression coefficients were similar to TEBC (n=86), but p-values were lower reflecting the increased statistical power achieved by combining samples. Overall, harmonization improves statistical power through increased sample sizes, however some additional variance may be introduced when combining samples which is not captured by harmonization.

To our knowledge this is the largest study to assess the relationship between gestational age at birth and neonatal dMRI metrics in healthy infants born >37.0 weeks. Average mean, axial and radial diffusivity and the histogram width for all tensor metrics were associated with GA at birth. Voxel-wise TBSS analyses revealed significant associations between lower GA at birth and lower FA, and higher MD, AD and RD throughout the white matter. This may represent altered white matter development at younger GA at birth within this sample of healthy term infants. In voxel-wise TBSS studies, lower FA and higher MD, AD and RD across the brain were associated with lower GA at birth (Broekman et al., 2014; Gale & Grant et al., 2022; Jin et al., 2019; Ou et al., 2017). Interestingly, we also identified a small region of the posterior/retrolenticular part of the left internal capsule where lower GA at birth was associated with higher FA. Post-hoc analyses revealed a significant positive correlation between FA in this region and postnatal age when adjusting for sex and gestational age at scan. It is possible this reflects increased maturation of sensorimotor white matter pathways, however this requires further investigation. Our results

are also in agreement with Blesa and colleagues who identified significant correlations between decreasing GA at birth and increased MD, AD and RD histogram widths in a cohort of preterm and term infants (Blesa et al., 2020). However, in contrast to Blesa and colleagues findings, a wider FA histogram in the white matter skeleton was associated with decreased GA at birth in this term-only sample. This difference is likely explained by differences in study populations as Blesa and colleagues included predominantly preterm infants. Increased mean, axial and radial diffusivity may reflect increased membrane permeability or altered oligodendrocyte proliferation (Pecheva et al., 2018), while wider FA histogram widths could represent changes to coordinate maturation patterns of white matter tracts postnatally (Ouyang et al., 2019) in infants born earlier in the term period, however this requires further investigation.

It is important to note that before harmonization we identified more widespread voxel-wise associations with GA at birth across DTI metrics in regions that were significantly associated with site. Therefore, failure to appropriately remove the effect of site may introduce inflated associations in neonatal data, particularly if site is associated either with the variable of interest or necessary covariates such as GA at scan. While introducing a linear covariate may mitigate differences between sites, it's important to note that within a GLM, the slope in each site is assumed to be equal. In our data, ComBat preserved differences in slope and data spread between sites when assessing the relationship between GA at birth and FA mean and histogram widths while linear covariates did not (as seen in Figure 4). Overall, this suggests harmonization with ComBat should be used to remove site effects rather than relying on adjusting for site with a linear covariate.

Finally, harmonizing dMRI from multiple centres to assemble large neonatal clinical cohorts has the potential to inform clinical practice. Clinical practice when treating populations in the

neonatal intensive care unit can vary across centers. By harmonizing dMRI data from multiple centers to remove the effect of site but preserve the effect of clinical factors, it is possible to assess the effect of different clinical strategies on white matter development in neonates as well as disentangle the interacting effects of clinical practice and non-modifiable clinical factors. However, in the context of large multi-centre clinical neuroimaging research studies, it is necessary to ensure the definition and recording of clinical factors such as neonatal sepsis or decision-making regarding delivery management is harmonized too. Future multi-centre clinical neuroimaging research studies should engage with clinical colleagues to identify clinical variables of importance and ensure definitions are consistent between sites.

Limitations and Future Directions

The aim of the study was to characterise the effect of ComBat on skeletonized DTI metrics in typically developing infants, therefore we did not assess the impact of data harmonization in a clinical population. Future studies should investigate the effect of ComBat on multi-centre neonatal dMRI case-control studies. In addition, further research is needed to assess the utility of novel harmonization methods such as deep learning techniques (Wada et al., 2023) for neonatal dMRI. In this study we chose to focus on harmonization of DTI metrics as these are widely used in neonatal dMRI research (Pecheva et al., 2018), and can be calculated from many conventional clinical dMRI acquisitions. However, other advanced measures of microstructure such as NODDI (Zhang et al., 2012) or fixel-based metrics (Raffelt et al., 2017) provide different insights into the developing brain and research assessing the impact of ComBat harmonization on these metrics in the neonatal brain is warranted. Finally, using dMRI to assess cortical microstructural development have reported alterations in neonates with congenital heart disease (Kelly et al., 2019) and those born prematurely (Ball et al., 2013; Galdi et al., 2024) and

correlations with clinically relevant factors (Kelly et al., 2019; Sullivan et al., 2023). Future studies are required to optimize pipelines for registration and preprocessing, as well as assessing the utility of comBat for multi-centre cortical diffusion research in neonates.

Conclusions

ComBat data harmonization removed the effect of site from skeletonised DTI metric means and histogram width in two cohorts of healthy neonates who underwent brain MRI using different scanning hardware and sequence parameters. Harmonization removed all voxel-wise differences from MD maps, however small differences remained in FA, AD and RD. In the large combined harmonized dataset, we were able to detect significant relationships between GA at birth and mean and histogram widths of DTI metrics. However, when comparing single site and multi-site harmonized datasets of equal sample sizes, harmonized data resulted in smaller standardised correlation coefficients for GA at birth. Overall, ComBat will enable the scale-up of neonatal neuroimaging studies to unprecedented sample sizes, which offers new horizons for biomarker discovery and validation, understanding upstream pathways to brain health and injury in early life, and the use of imaging for investigating the efficacy of neuroprotective therapies.

5. Acknowledgements and Funding

This research was funded by the Harold Hyam Wingate Foundation and Medical Research Council UK (MR/V002465/1). JPB acknowledges funding from the UKRI MRC programme grant (MR/X003434/1) and Theirworld (<http://www.theirworld.org/>). The Developing Human Connectome Project was funded by the European Research Council under the European Union's Seventh Framework Program (FP7/20072013)/European Research Council grant agreement no.

319456. This research was supported by core funding from the Wellcome/EPSRC Centre for Medical Engineering (WT203148/Z/16/Z) and by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- Andersson, J. L. R., Graham, M. S., Drobnyak, I., Zhang, H., Filippini, N., & Bastiani, M. (2017). Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *NeuroImage*, *152*, 450–466. <https://doi.org/10.1016/j.neuroimage.2017.02.085>
- Andersson, J. L. R., Graham, M. S., Zsoldos, E., & Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, *141*, 556–572. <https://doi.org/10.1016/j.neuroimage.2016.06.058>
- Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. [https://doi.org/10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7)
- Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>
- Azzopardi, D., Robertson, N. J., Bainbridge, A., Cady, E., Charles-Edwards, G., Deierl, A., Fagiolo, G., Franks, N. P., Griffiths, J., Hajnal, J., Juszczak, E., Kapetanakis, B., Linsell, L.,

- Maze, M., Omar, O., Strohm, B., Tusor, N., & Edwards, A. D. (2016). Moderate hypothermia within 6 h of birth plus inhaled xenon versus moderate hypothermia alone after birth asphyxia (TOBY-Xe): a proof-of-concept, open-label, randomised controlled trial. *The Lancet Neurology*, *15*(2), 145–153. [https://doi.org/10.1016/S1474-4422\(15\)00347-6](https://doi.org/10.1016/S1474-4422(15)00347-6)
- Barnett, M. L., Tusor, N., Ball, G., Chew, A., Falconer, S., Aljabar, P., Kimpton, J. A., Kennea, N., Rutherford, M., David Edwards, A., & Counsell, S. J. (2018). Exploring the multiple-hit hypothesis of preterm white matter damage using diffusion MRI. *NeuroImage. Clinical*, *17*, 596–606. <https://doi.org/10.1016/j.nicl.2017.11.017>
- Basser, P. J., Mattiello, J., & Lebihan, D. (1994). Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B*, *103*(3), 247–254. <https://doi.org/10.1006/jmrb.1994.1037>
- Baykara, E., Gesierich, B., Adam, R., Tuladhar, A. M., Biesbroek, J. M., Koek, H. L., Ropele, S., Jouvent, E., Chabriat, H., Ertl-Wagner, B., Ewers, M., Schmidt, R., de Leeuw, F., Biessels, G. J., Dichgans, M., & Duering, M. (2016). A Novel Imaging Marker for Small Vessel Disease Based on Skeletonization of White Matter Tracts and Diffusion Histograms. *Annals of Neurology*, *80*(4), 581–592. <https://doi.org/10.1002/ana.24758>
- Ball G, Counsell SJ, Anjari M, Merchant N, Arichi T, Doria V, Rutherford MA, Edwards AD, Rueckert D, Boardman JP. An optimised tract-based spatial statistics protocol for neonates: applications to prematurity and chronic lung disease. (2010) *Neuroimage*, *53*(1):94-102. [10.1016/j.neuroimage.2010.05.055](https://doi.org/10.1016/j.neuroimage.2010.05.055). Blesa, M., Galdi, P., Sullivan, G., Wheeler, E. N., Stoye, D. Q., Lamb, G. J., Quigley, A. J., Thrippleton, M. J., Bastin, M. E., & Boardman, J. P. (2020). Peak Width of Skeletonized Water Diffusion MRI in the Neonatal Brain. *Frontiers in Neurology*, *11*. <https://doi.org/10.3389/fneur.2020.00235>

- Blesa, M., Serag, A., Wilkinson, A. G., Anblagan, D., Telford, E. J., Pataky, R., Sparrow, S. A., Macnaught, G., Semple, S. I., Bastin, M. E., & Boardman, J. P. (2016). Parcellation of the Healthy Neonatal Brain into 107 Regions Using Atlas Propagation through Intermediate Time Points in Childhood. *Frontiers in Neuroscience*, *10*. <https://doi.org/10.3389/fnins.2016.00220>
- Blesa, M., Sullivan, G., Anblagan, D., Telford, E. J., Quigley, A. J., Sparrow, S. A., Serag, A., Semple, S. I., Bastin, M. E., & Boardman, J. P. (2019). Early breast milk exposure modifies brain connectivity in preterm infants. *NeuroImage*, *184*, 431–439. <https://doi.org/10.1016/j.neuroimage.2018.09.045>
- Boardman, J. P., & Counsell, S. J. (2020). Invited Review: Factors associated with atypical brain development in preterm infants: insights from magnetic resonance imaging. *Neuropathology and Applied Neurobiology*, *46*(5), 413–421. <https://doi.org/10.1111/nan.12589>
- Boardman, J. P., Hall, J., Thrippleton, M. J., Reynolds, R. M., Bogaert, D., Davidson, D. J., Schwarze, J., Drake, A. J., Chandran, S., Bastin, M. E., & Fletcher-Watson, S. (2020). Impact of preterm birth on brain development and long-term outcome: protocol for a cohort study in Scotland. *BMJ Open*, *10*(3), e035854. <https://doi.org/10.1136/bmjopen-2019-035854>
- Broekman, B. F. P., Wang, C., Li, Y., Rifkin-Graboi, A., Saw, S. M., Chong, Y.-S., Kwek, K., Gluckman, P. D., Fortier, M. V., Meaney, M. J., & Qiu, A. (2014a). Gestational Age and Neonatal Brain Microstructure in Term Born Infants: A Birth Cohort Study. *PLoS ONE*, *9*(12), e115229. <https://doi.org/10.1371/journal.pone.0115229>

- Broekman, B. F. P., Wang, C., Li, Y., Rifkin-Graboi, A., Saw, S. M., Chong, Y.-S., Kwek, K., Gluckman, P. D., Fortier, M. V., Meaney, M. J., & Qiu, A. (2014b). Gestational Age and Neonatal Brain Microstructure in Term Born Infants: A Birth Cohort Study. *PLoS ONE*, 9(12), e115229. <https://doi.org/10.1371/journal.pone.0115229>
- Christiaens, D., Cordero-Grande, L., Pietsch, M., Hutter, J., Price, A. N., Hughes, E. J., Vecchiato, K., Deprez, M., Edwards, A. D., Hajnal, J. V., & Tournier, J.-D. (2021). Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI. *NeuroImage*, 225, 117437. <https://doi.org/10.1016/j.neuroimage.2020.117437>
- Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A. N., & Hajnal, J. V. (2019). Complex diffusion-weighted image estimation via matrix recovery under general noise models. *NeuroImage*, 200, 391–404. <https://doi.org/10.1016/j.neuroimage.2019.06.039>
- Demers, C. H., Bagonis, M. M., Al-Ali, K., Garcia, S. E., Styner, M. A., Gilmore, J. H., Hoffman, M. C., Hankin, B. L., & Davis, E. P. (2021). Exposure to prenatal maternal distress and infant white matter neurodevelopment. *Development and Psychopathology*, 33(5), 1526–1538. <https://doi.org/10.1017/S0954579421000742>
- Dibble, M., Ang, J. Z., Mariga, L., Molloy, E. J., & Bokde, A. L. W. (2021). Diffusion Tensor Imaging in Very Preterm, Moderate-Late Preterm and Term-Born Neonates: A Systematic Review. *The Journal of Pediatrics*, 232, 48-58.e3. <https://doi.org/10.1016/j.jpeds.2021.01.008>
- Dibble, M., O’Dea, M. I., Hurley, T., Byrne, A., Colleran, G., Molloy, E. J., & Bokde, A. L. W. (2020). Diffusion tensor imaging in neonatal encephalopathy: a systematic review. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 105(5), 480–488. <https://doi.org/10.1136/archdischild-2019-318025>

- Dimitrova, R., Pietsch, M., Christiaens, D., Ciarrusta, J., Wolfers, T., Batalle, D., Hughes, E., Hutter, J., Cordero-Grande, L., Price, A. N., Chew, A., Falconer, S., Vecchiato, K., Steinweg, J. K., Carney, O., Rutherford, M. A., Tournier, J.-D., Counsell, S. J., Marquand, A. F., ... O'Muircheartaigh, J. (2020). Heterogeneity in Brain Microstructural Development Following Preterm Birth. *Cerebral Cortex*, 30(9), 4800–4810. <https://doi.org/10.1093/cercor/bhaa069>
- Dubois, J., Alison, M., Counsell, S. J., Hertz-Pannier, L., Hüppi, P. S., & Benders, M. J. N. L. (2021). MRI of the Neonatal Brain: A Review of Methodological Challenges and Neuroscientific Advances. *Journal of Magnetic Resonance Imaging*, 53(5), 1318–1343. <https://doi.org/10.1002/jmri.27192>
- Edwards, A. D., Rueckert, D., Smith, S. M., Abo Seada, S., Alansary, A., Almalbis, J., Allsop, J., Andersson, J., Arichi, T., Arulkumaran, S., Bastiani, M., Batalle, D., Baxter, L., Bozek, J., Braithwaite, E., Brandon, J., Carney, O., Chew, A., Christiaens, D., ... Hajnal, J. V. (2022). The Developing Human Connectome Project Neonatal Data Release. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.886772>
- Feng, K., Rowell, A. C., Andres, A., Bellando, B. J., Lou, X., Glasier, C. M., Ramakrishnaiah, R. H., Badger, T. M., & Ou, X. (2019). Diffusion Tensor MRI of White Matter of Healthy Full-term Newborns: Relationship to Neurodevelopmental Outcomes. *Radiology*, 292(1), 179–187. <https://doi.org/10.1148/radiol.2019182564>
- Feng, L., Li, H., Oishi, K., Mishra, V., Song, L., Peng, Q., Ouyang, M., Wang, J., Slinger, M., Jeon, T., Lee, L., Heyne, R., Chalak, L., Peng, Y., Liu, S., & Huang, H. (2019). Age-specific gray and white matter DTI atlas for human brain at 33, 36 and 39 postmenstrual weeks. *NeuroImage*, 185, 685–698. <https://doi.org/10.1016/j.neuroimage.2018.06.069>

- 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., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. <https://doi.org/10.1016/j.neuroimage.2017.08.047>
- Galdi, P., Blesa, M., Sullivan, G., Lamb, G. J., Stoye, D. Q., Quigley, A. J., Thrippleton, M. J., Bastin, M. E., & Boardman, J. P. (2018). *Neonatal Morphometric Similarity Networks Predict Atypical Brain Development Associated with Preterm Birth* (pp. 47–57). https://doi.org/10.1007/978-3-030-00755-3_6
- Galdi, P., Cabez, M. B., Farrugia, C., Vaher, K., Williams, L. Z. J., Sullivan, G., Stoye, D. Q., Quigley, A. J., Makropoulos, A., Thrippleton, M. J., Bastin, M. E., Richardson, H., Whalley, H., Edwards, A. D., Bajada, C. J., Robinson, E. C., & Boardman, J. P. (2024). Feature similarity gradients detect alterations in the neonatal cortex associated with preterm birth. *Human Brain Mapping*, *45*(4). <https://doi.org/10.1002/hbm.26660>
- Gale Grant, O., Fenn Moltu, S., França, L. G. S., Dimitrova, R., Christiaens, D., Cordero Grande, L., Chew, A., Falconer, S., Harper, N., Price, A. N., Hutter, J., Hughes, E., O’Muircheartaigh, J., Rutherford, M., Counsell, S. J., Rueckert, D., Nosarti, C., Hajnal, J. V., McAlonan, G., ... Batalle, D. (2022). Effects of gestational age at birth on perinatal structural brain development in healthy term-born babies. *Human Brain Mapping*, *43*(5), 1577–1589. <https://doi.org/10.1002/hbm.25743>
- Howell, B. R., Styner, M. A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., Li, G., Gilmore, J. H., Piven, J., Smith, J. K., Shen, D., Ugurbil, K., Zhu, H., Lin, W., & Elison, J. T. (2019). The UNC/UMN Baby Connectome Project (BCP):

- An overview of the study design and protocol development. *NeuroImage*, 185, 891–905.
<https://doi.org/10.1016/j.neuroimage.2018.03.049>
- Hua, J., Sun, J., Cao, Z., Dai, X., Lin, S., Guo, J., Gu, G., & Du, W. (2019). Differentiating the cognitive development of early-term births in infants and toddlers: a cross-sectional study in China. *BMJ Open*, 9(4), e025275. <https://doi.org/10.1136/bmjopen-2018-025275>
- Hughes, E. J., Winchman, T., Padormo, F., Teixeira, R., Wurie, J., Sharma, M., Fox, M., Hutter, J., Cordero-Grande, L., Price, A. N., Allsop, J., Bueno-Conde, J., Tusor, N., Arichi, T., Edwards, A. D., Rutherford, M. A., Counsell, S. J., & Hajnal, J. V. (2017). A dedicated neonatal brain imaging system. *Magnetic Resonance in Medicine*, 78(2), 794–804.
<https://doi.org/10.1002/mrm.26462>
- Hutter, J., Tournier, J. D., Price, A. N., Cordero-Grande, L., Hughes, E. J., Malik, S., Steinweg, J., Bastiani, M., Sotiropoulos, S. N., Jbabdi, S., Andersson, J., Edwards, A. D., & Hajnal, J. V. (2018). Time-efficient and flexible design of optimized multishell HARDI diffusion. *Magnetic Resonance in Medicine*, 79(3), 1276–1292. <https://doi.org/10.1002/mrm.26765>
- Jin, C., Li, Y., Li, X., Liu, C., Wang, M., Cheng, Y., Zheng, J., & Yang, J. (2019). Associations of gestational age and birth anthropometric indicators with brain white matter maturation in full-term neonates. *Human Brain Mapping*, 40(12), 3620–3630.
<https://doi.org/10.1002/hbm.24620>
- 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>
- Karmacharya, S., Gagoski, B., Ning, L., Vyas, R., Cheng, H. H., Soul, J., Newberger, J. W., Shenton, M. E., Rathi, Y., & Grant, P. E. (2018). Advanced diffusion imaging for assessing

normal white matter development in neonates and characterizing aberrant development in congenital heart disease. *NeuroImage. Clinical*, 19, 360–373.

<https://doi.org/10.1016/j.nicl.2018.04.032>

Kellner, E., Dhital, B., Kiselev, V. G., & Reisert, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine*, 76(5), 1574–1581.

<https://doi.org/10.1002/mrm.26054>

Kelly, C. J., Christiaens, D., Batalle, D., Makropoulos, A., Cordero-Grande, L., Steinweg, J. K., O’Muircheartaigh, J., Khan, H., Lee, G., Victor, S., Alexander, D. C., Zhang, H., Simpson, J., Hajnal, J. V., Edwards, A. D., Rutherford, M. A., & Counsell, S. J. (2019). Abnormal Microstructural Development of the Cerebral Cortex in Neonates With Congenital Heart Disease Is Associated With Impaired Cerebral Oxygen Delivery. *Journal of the American Heart Association*, 8(5), e009893. <https://doi.org/10.1161/JAHA.118.009893>

Korom, M., Camacho, M. C., Filippi, C. A., Licandro, R., Moore, L. A., Dufford, A., Zöllei, L., Graham, A. M., Spann, M., Howell, B., Shultz, S., & Scheinost, D. (2022). Dear reviewers: Responses to common reviewer critiques about infant neuroimaging studies. *Developmental Cognitive Neuroscience*, 53, 101055. <https://doi.org/10.1016/j.dcn.2021.101055>

Lautarescu, A., Bonthron, A. F., Pietsch, M., Batalle, D., Cordero-Grande, L., Tournier, J.-D., Christiaens, D., Hajnal, J. V., Chew, A., Falconer, S., Nosarti, C., Victor, S., Craig, M. C., Edwards, A. D., & Counsell, S. J. (2022). Maternal depressive symptoms, neonatal white matter, and toddler social-emotional development. *Translational Psychiatry*, 12(1), 323. <https://doi.org/10.1038/s41398-022-02073-y>

Lautarescu, A., Pecheva, D., Nosarti, C., Nihouarn, J., Zhang, H., Victor, S., Craig, M., Edwards, A. D., & Counsell, S. J. (2019). Maternal Prenatal Stress Is Associated With Altered

Uncinate Fasciculus Microstructure in Premature Neonates. *Biological Psychiatry*.

<https://doi.org/10.1016/j.biopsych.2019.08.010>

Law, J. B., Comstock, B. A., Richards, T. L., Traudt, C. M., Wood, T. R., Mayock, D. E., Heagerty, P. J., & Juul, S. E. (2021). Diffusion Tensor Imaging Changes Do Not Affect Long-Term Neurodevelopment following Early Erythropoietin among Extremely Preterm Infants in the Preterm Erythropoietin Neuroprotection Trial. *Brain Sciences*, *11*(10), 1360.

<https://doi.org/10.3390/brainsci11101360>

Le Bihan, D., Mangin, J., Poupon, C., Clark, C. A., Pappata, S., Molko, N., & Chabriet, H. (2001). Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, *13*(4), 534–546. <https://doi.org/10.1002/jmri.1076>

Lean, R. E., Smyser, C. D., Brady, R. G., Triplett, R. L., Kaplan, S., Kenley, J. K., Shimony, J. S., Smyser, T. A., Miller, J. P., Barch, D. M., Luby, J. L., Warner, B. B., & Rogers, C. E. (2022). Prenatal exposure to maternal social disadvantage and psychosocial stress and neonatal white matter connectivity at birth. *Proceedings of the National Academy of Sciences*, *119*(42). <https://doi.org/10.1073/pnas.2204135119>

Lourenço, V. M., Pires, A. M., & Kirst, M. (2011). Robust linear regression methods in association studies. *Bioinformatics (Oxford, England)*, *27*(6), 815–821.

<https://doi.org/10.1093/bioinformatics/btr006>

Meyers, B., Lee, V. K., Dennis, L., Wallace, J., Schmithorst, V., Votava-Smith, J. K., Rajagopalan, V., Herrup, E., Baust, T., Tran, N. N., Hunter, J. V., Licht, D. J., Gaynor, J. W., Andropoulos, D. B., Panigrahy, A., & Ceschin, R. (2022). Harmonization of multi-center diffusion tensor tractography in neonates with congenital heart disease: Optimizing

- post-processing and application of ComBat. *Neuroimage: Reports*, 2(3), 100114.
<https://doi.org/10.1016/j.ynirp.2022.100114>
- Miller, S. P., McQuillen, P. S., Hamrick, S., Xu, D., Glidden, D. V, Charlton, N., Karl, T., Azakie, A., Ferriero, D. M., Barkovich, A. J., & Vigneron, D. B. (2007). Abnormal brain development in newborns with congenital heart disease. *The New England Journal of Medicine*, 357(19), 1928–1938. <https://doi.org/10.1056/NEJMoa067393>
- Mirzaalian, H., de Pierrefeu, A., Savadjiev, P., Pasternak, O., Bouix, S., Kubicki, M., Westin, C.-F., Shenton, M. E., & Rathi, Y. (2015). Harmonizing Diffusion MRI Data Across Multiple Sites and Scanners (pp. 12–19). https://doi.org/10.1007/978-3-319-24553-9_2
- Moltu, S. J., Nordvik, T., Rossholt, M. E., Wendel, K., Chawla, M., Server, A., Gunnarsdottir, G., Pripp, A. H., Domellöf, M., Bratlie, M., Aas, M., Hüppi, P. S., Lapillonne, A., Beyer, M. K., Stiris, T., Maximov, I. I., Geier, O., & Pfeiffer, H. (2024). Arachidonic and docosahexaenoic acid supplementation and brain maturation in preterm infants; a double blind RCT. *Clinical Nutrition*, 43(1), 176–186. <https://doi.org/10.1016/j.clnu.2023.11.037>
- Mulkey, S. B., Ou, X., Ramakrishnaiah, R. H., Glasier, C. M., Swearingen, C. J., Melguizo, M. S., Yap, V. L., Schmitz, M. L., & Bhutta, A. T. (2014). White matter injury in newborns with congenital heart disease: a diffusion tensor imaging study. *Pediatric Neurology*, 51(3), 377–383. <https://doi.org/10.1016/j.pediatrneurol.2014.04.008>
- O’Gorman, R. L., Bucher, H. U., Held, U., Koller, B. M., Hüppi, P. S., & Hagmann, C. F. (2015). Tract-based spatial statistics to assess the neuroprotective effect of early erythropoietin on white matter development in preterm infants. *Brain*, 138(2), 388–397. <https://doi.org/10.1093/brain/awu363>

- Ou, X., Glasier, C. M., Ramakrishnaiah, R. H., Kanfi, A., Rowell, A. C., Pivik, R. T., Andres, A., Cleves, M. A., & Badger, T. M. (2017a). Gestational Age at Birth and Brain White Matter Development in Term-Born Infants and Children. *American Journal of Neuroradiology*, *38*(12), 2373–2379. <https://doi.org/10.3174/ajnr.A5408>
- Parikh, M. N., Chen, M., Braimah, A., Kline, J., McNally, K., Logan, J. W., Tamm, L., Yeates, K. O., Yuan, W., He, L., & Parikh, N. A. (2021). Diffusion MRI Microstructural Abnormalities at Term-Equivalent Age Are Associated with Neurodevelopmental Outcomes at 3 Years of Age in Very Preterm Infants. *American Journal of Neuroradiology*, *42*(8), 1535–1542. <https://doi.org/10.3174/ajnr.A7135>
- Pecheva, D., Kelly, C., Kimpton, J., Bonthron, A., Batalle, D., Zhang, H., & Counsell, S. J. (2018). Recent advances in diffusion neuroimaging: applications in the developing preterm brain. *FI000Research*, *7*, 1326. <https://doi.org/10.12688/f1000research.15073.1>
- Peña, E. A., & Slate, E. H. (2006). Global Validation of Linear Model Assumptions. *Journal of the American Statistical Association*, *101*(473), 341–354. <https://doi.org/10.1198/016214505000000637>
- Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I. M., Satterthwaite, T. D., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Wolf, D. H., ... Davatzikos, C. (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>
- Poppe, T., Thompson, B., Boardman, J. P., Bastin, M. E., Alsweiler, J., Deib, G., Harding, J. E., & Crowther, C. A. (2022). Effect of antenatal magnesium sulphate on MRI biomarkers of

- white matter development at term equivalent age: The MagNUM Study. *EBioMedicine*, 78, 103923. <https://doi.org/10.1016/j.ebiom.2022.103923>
- Porter, E. J., Counsell, S. J., Edwards, A. D., Allsop, J., & Azzopardi, D. (2010). Tract-Based Spatial Statistics of Magnetic Resonance Images to Assess Disease and Treatment Effects in Perinatal Asphyxial Encephalopathy. *Pediatric Research*, 68(3), 205–209. <https://doi.org/10.1203/PDR.0b013e3181e9f1ba>
- Raffelt, D. A., Tournier, J.-D. D., Smith, R. E., Vaughan, D. N., Jackson, G., Ridgway, G. R., & Connelly, A. (2017). Investigating white matter fibre density and morphology using fixel-based analysis. *Neuroimage*, 144(Pt A), 58–73. <https://doi.org/10.1016/j.neuroimage.2016.09.029>
- Rose, O., Blanco, E., Martinez, S. M., Sim, E. K., Castillo, M., Lozoff, B., Vaucher, Y. E., & Gahagan, S. (2013). Developmental Scores at 1 Year With Increasing Gestational Age, 37–41 Weeks. *Pediatrics*, 131(5), e1475–e1481. <https://doi.org/10.1542/peds.2012-3215>
- Salvan, P., Tournier, J. D., Batalle, D., Falconer, S., Chew, A., Kennea, N., Aljabar, P., Dehaene, Lambertz, G., Arichi, T., Edwards, A. D., & Counsell, S. J. (2017). Language ability in preterm children is associated with arcuate fasciculi microstructure at term. *Human Brain Mapping*, 38(8), 3836–3847. <https://doi.org/10.1002/hbm.23632>
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. J. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4), 1487–1505. <https://doi.org/10.1016/j.neuroimage.2006.02.024>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders,

- J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1, S208-19. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Soh, S.-E., Tint, M. T., Gluckman, P. D., Godfrey, K. M., Rifkin-Graboi, A., Chan, Y. H., Stunkel, W., Holbrook, J. D., Kwek, K., Chong, Y.-S., Saw, S. M., Sheppard, A., Chinnadurai, A., Ferguson-Smith, A., Goh, A. E. N., Biswas, A., Chia, A., Leutscher-Broekman, B., Shuter, B., ... Venkatesh, S. K. (2014). Cohort Profile: Growing Up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study. *International Journal of Epidemiology*, 43(5), 1401–1409. <https://doi.org/10.1093/ije/dyt125>
- Sonuga-Barke, E. J. S. (2023). Editorial: “Safety in numbers”? Big data discovery strategies in neurodevelopmental science – contributions and caveats. *Journal of Child Psychology and Psychiatry*, 64(1), 1–3. <https://doi.org/10.1111/jcpp.13723>
- Stoye, D. Q., Blesa, M., Sullivan, G., Galdi, P., Lamb, G. J., Black, G. S., Quigley, A. J., Thrippleton, M. J., Bastin, M. E., Reynolds, R. M., & Boardman, J. P. (2020). Maternal cortisol is associated with neonatal amygdala microstructure and connectivity in a sexually dimorphic manner. *ELife*, 9. <https://doi.org/10.7554/eLife.60729>
- Systemic inflammation: Sullivan G, Galdi P, Cabez MB, Borbye-Lorenzen N, Stoye DQ, Lamb GJ, Evans MJ, Quigley AJ, Thrippleton MJ, Skogstrand K, Chandran S, Bastin ME, Boardman JP. (2020) Interleukin-8 dysregulation is implicated in brain dysmaturation following preterm birth. *Brain Behav Immun*. 90:311-318. 10.1016/j.bbi.2020.09.007.
- Sullivan, G., Vaher, K., Blesa, M., Galdi, P., Stoye, D. Q., Quigley, A. J., Thrippleton, M. J., Norrie, J., Bastin, M. E., & Boardman, J. P. (2023). Breast Milk Exposure is Associated

With Cortical Maturation in Preterm Infants. *Annals of Neurology*, 93(3), 591–603.

<https://doi.org/10.1002/ana.26559>

Tusor, N., Wusthoff, C., Smeets, N., Merchant, N., Arichi, T., Allsop, J. M., Cowan, F. M., Azzopardi, D., Edwards, A. D., & Counsell, S. J. (2012). Prediction of neurodevelopmental outcome after hypoxic–ischemic encephalopathy treated with hypothermia by diffusion tensor imaging analyzed using tract-based spatial statistics. *Pediatric Research*, 72(1), 63–69. <https://doi.org/10.1038/pr.2012.40>

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>

Ullman, H., Spencer-Smith, M., Thompson, D. K., Doyle, L. W., Inder, T. E., Anderson, P. J., & Klingberg, T. (2015). Neonatal MRI is associated with future cognition and academic achievement in preterm children. *Brain*, 138(11), 3251–3262. <https://doi.org/10.1093/brain/awv244>

Vaher, K., Galdi, P., Blesa Cabeza, M., Sullivan, G., Stoye, D. Q., Quigley, A. J., Thrippleton, M. J., Bogaert, D., Bastin, M. E., Cox, S. R., & Boardman, J. P. (2022). General factors of white matter microstructure from DTI and NODDI in the developing brain. *NeuroImage*, 254, 119169. <https://doi.org/10.1016/j.neuroimage.2022.119169>

Veraart, J., Fieremans, E., & Novikov, D. S. (2016). Diffusion MRI noise mapping using random matrix theory. *Magnetic Resonance in Medicine*, 76(5), 1582–1593. <https://doi.org/10.1002/mrm.26059>

Wada, A., Akashi, T., Hagiwara, A., Nishizawa, M., Shimoji, K., Kikuta, J., Maekawa, T., Sano, K., Kamagata, K., Nakanishi, A., & Aoki, S. (2023). Deep Learning-Driven

Transformation: A Novel Approach for Mitigating Batch Effects in Diffusion MRI Beyond Traditional Harmonization. *Journal of Magnetic Resonance Imaging*.
<https://doi.org/10.1002/jmri.29088>

Wang, Y., Gupta, A., Liu, Z., Zhang, H., Escolar, M. L., Gilmore, J. H., Gouttard, S., Fillard, P., Maltbie, E., Gerig, G., & Styner, M. (2011). DTI registration in atlas based fiber analysis of infantile Krabbe disease. *NeuroImage*, 55(4), 1577–1586.
<https://doi.org/10.1016/j.neuroimage.2011.01.038>

Ward, P., Counsell, S., Allsop, J., Cowan, F., Shen, Y., Edwards, D., & Rutherford, M. (2006). Reduced Fractional Anisotropy on Diffusion Tensor Magnetic Resonance Imaging After Hypoxic-Ischemic Encephalopathy. *Pediatrics*, 117(4), e619–e630.
<https://doi.org/10.1542/peds.2005-0545>

Wheater ENW, Galdi P, McCartney DL, Blesa M, Sullivan G, Stoye DQ, Lamb G, Sparrow S, Murphy L, Wrobel N, Quigley AJ, Semple S, Thrippleton MJ, Wardlaw JM, Bastin ME, Marioni RE, Cox SR, Boardman JP (2022). DNA methylation in relation to gestational age and brain dysmaturation in preterm infants. *Brain Commun*. 4(2):fcac056. 10.1093/braincomms/fcac056.

Zhang, H., Schneider, T., Wheeler-Kingshott, C. a, & Alexander, D. C. (2012). NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 61(4), 1000–1016. <https://doi.org/10.1016/j.neuroimage.2012.03.072>

Zhang, H., Yushkevich, P., Alexander, D., & Gee, J. (2006). Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*, 10(5), 764–785. <https://doi.org/10.1016/j.media.2006.06.004>

Figures

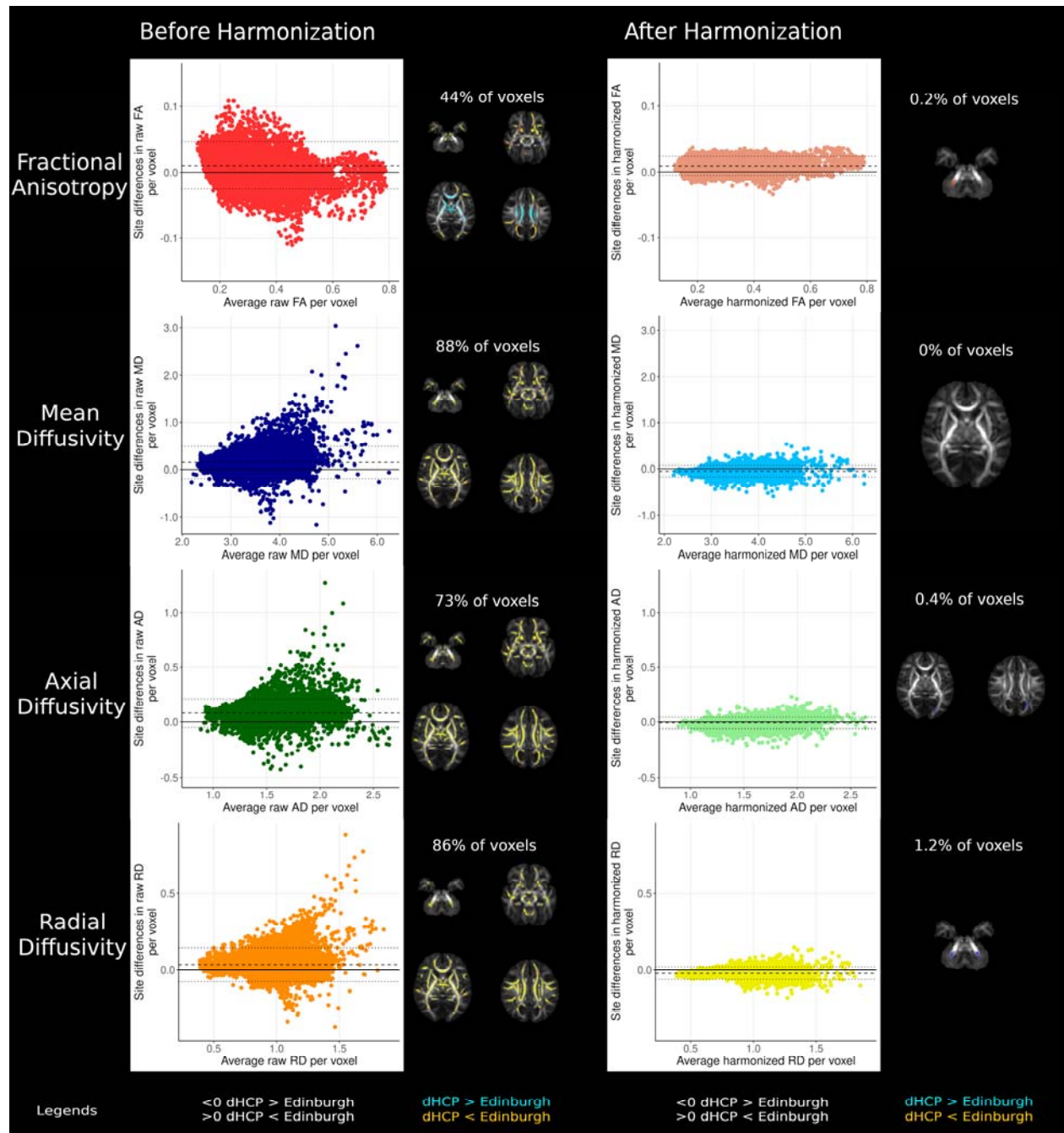


Figure 1. Voxel-wise differences between sites before and after harmonization. Mean difference plots with dashed lines representing mean difference and dotted lines representing 95% confidence intervals without adjusting for gestational age at birth, gestational age at scan and sex.

Voxel percentages and results displayed on mean FA images represent voxels significantly different between sites after adjusting for sex, gestational age at birth and gestational age at scan.

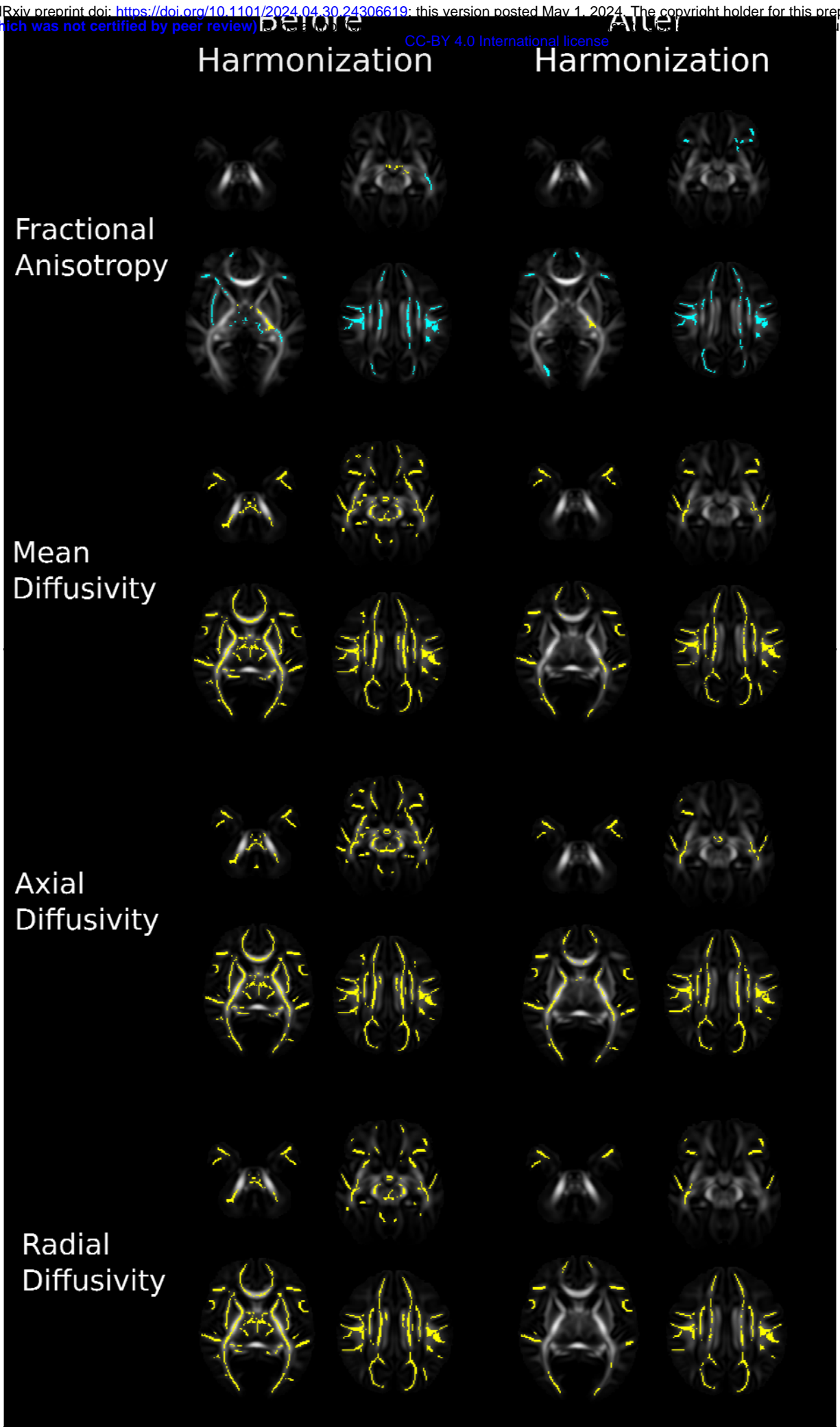


Figure 2. Voxel-wise correlations with GA at birth across the whole sample before and after harmonization. Results displayed on mean FA images represent voxels significantly positively (blue) and negatively (yellow) associated with gestational age at birth adjusting for sex and gestational age at scan.

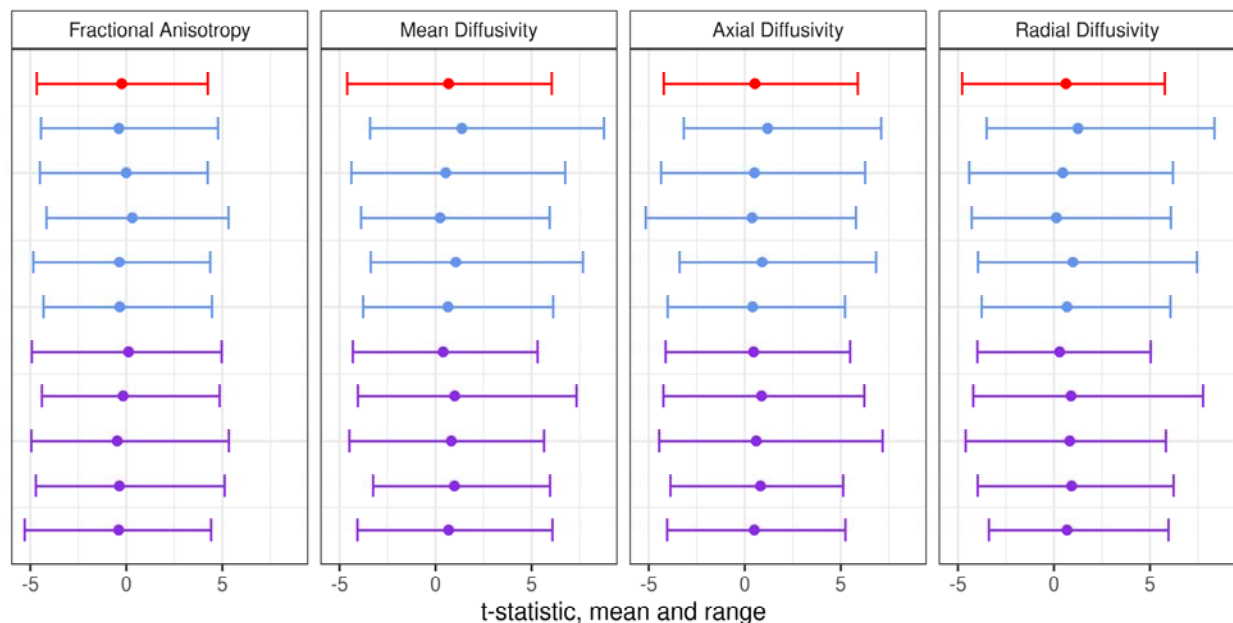


Figure 3. Mean and range of t-statistics for voxel-wise relationships between DTI metrics and GA at birth in Theirworld Edinburgh Birth Cohort (n=86; red), five subsamples of developing human connectome project (n=86 in each; blue) and five subsamples of combined cohort with ComBat harmonization (n=86 in each; purple).

Figure 4. Voxels significantly associated with GA at birth across five subsets of dHCP and harmonized dHCP and TEBC. Results displayed on mean FA images represent number of times a voxel was associated with GA at birth in each subset (0-5 times) when adjusting for sex and gestational age at scan.

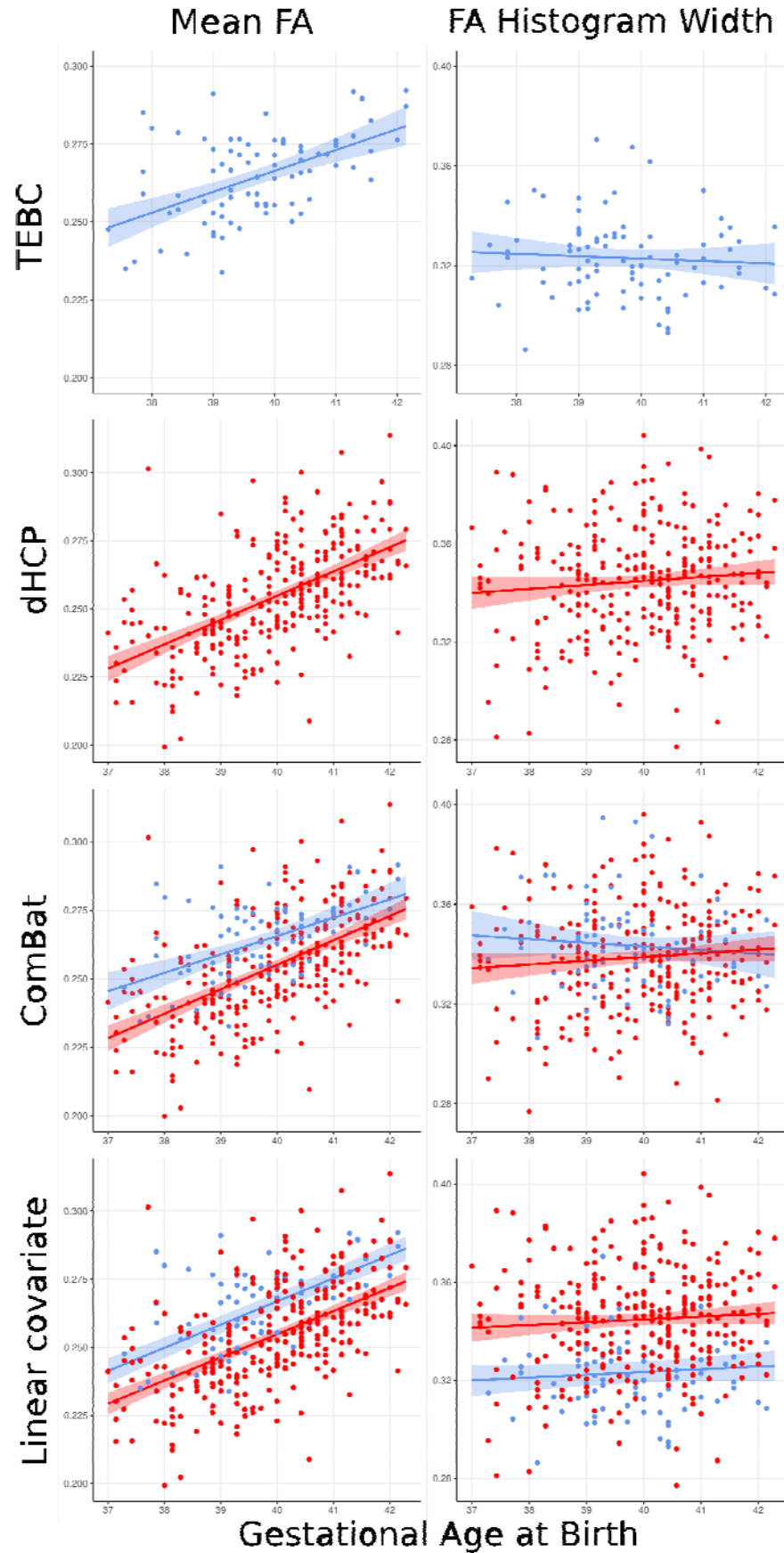


Figure 5 Relationship between GA at birth in Theirworld Edinburgh Birth Cohort (TEBC), developing human connectome project (dHCP) and combined cohort with ComBat harmonization and site as a linear covariate. Blue- TEBC; Red- dHCP.

Tables

Table 1. Demographics and Characteristics of dHCP and TEBC			
	Developing Human Connectome Project (n=287)	Theirworld Edinburgh Birth Cohort (n=86)	Differences between sites
Gestational age at birth, median (IQR)	40.14 (39.00-40.86)	39.71 (39.00-40.43)	p=0.076
GA at scan, median (IQR)	40.86 (39.81-42.14)	42.00 (41.29-43.00)	p<0.001
Female, n (%)	134 (46.7)	40 (46.5)	X ² =0 p=1.0
results in bold are significant			

Table 2. Mean and histogram widths for each site before and after harmonization

Diffusion metric	Measure, mean (SD)	Before Harmonization		After Harmonization	
		TEBC (n=86)	dHCP (n=287)	TEBC (n=86)	dHCP (n=287)
Fractional anisotropy	Mean	0.265 (0.014)	0.254 (0.019)	0.265 (0.014)	0.255 (0.019)
	Histogram width	0.323 (0.016)	0.345 (0.022)	0.343 (0.017)	0.338 (0.022)
Mean Diffusivity	Mean	3.79 (0.114)	3.65 (0.156)	3.64 (0.119)	3.69 (0.155)
	Histogram Width	1.49 (0.160)	1.57 (0.188)	1.54 (0.168)	1.55 (0.186)
Axial Diffusivity	Mean	1.63 (0.031)	1.55 (0.041)	1.56 (0.031)	1.57 (0.041)
	Histogram Width	0.702 (0.037)	0.656 (0.048)	0.668 (0.034)	0.660 (0.048)
Radial Diffusivity	Mean	1.08 (0.043)	1.05 (0.059)	1.04 (0.044)	1.06 (0.058)
	Histogram Width	0.603 (0.058)	0.649 (0.066)	0.634 (0.061)	0.637 (0.065)

Table 3. Effect of site on dMRI metrics before and after harmonization			
Diffusion metric	Measure	Effect of site before harmonization	Effect of site after harmonization
Fractional anisotropy	Mean	β (SE)= -0.056 (0.037) p=0.128 ^a	β (SE)= -0.032 (0.037) p=0.378 ^a
	Histogram width	β (SE)= -0.568 (0.048) p<0.001^a	β (SE)= 0.063 (0.054) p=0.243
Mean Diffusivity	Mean	β (SE)= -0.485 (0.040) p<0.001	β (SE)= 0.039 (0.043) p=0.360
	Histogram Width	β (SE)= 0.171 (0.052) p=0.001	β (SE)= 0.014 (0.053) p=0.788
Axial Diffusivity	Mean	β (SE)= -0.684(0.037) p<0.001	β (SE)= 0.048 (0.049) p=0.328
	Histogram Width	β (SE)= -0.376 (0.052) p<0.001^a	β (SE)=-0.006 (0.057) p=0.921 ^a
Radial Diffusivity	Mean	β (SE)= -0.370(0.040) p<0.001	β (SE)= 0.035 (0.041) p=0.389
	Histogram Width	β (SE)= 0.284 (0.049) p<0.001	β (SE)= 0.020 (0.051) p=0.692
adjusted for GA at birth, GA at scan and sex; results in bold are significant, ^a Robust regression			

Diffusion metric	Measure	Gestational age at birth in TEBC (n=86)	Gestational age at birth in dHCP (n=287)	Gestational age at birth with no correction (n=373)	Gestational age at birth with site as a covariate (n=373)	Gestational age at birth after harmonization (n=373)
Fractional Anisotropy	Mean	β (SE)= 0.084 (0.092) p=0.363	β (SE)= 0.088 (0.057) p=0.125	β (SE)= 0.068 (0.044) p=0.127 ^a	β (SE)= -0.094 (0.047) p=0.047^a	β (SE)= 0.08 (0.044) p=0.072 ^a
	Histogram width	β (SE)= -0.383 (0.130) p=0.004	β (SE)= -0.248 (0.081) p=0.002	β (SE)= -0.001 (0.068) p=0.985	β (SE)= -0.258 (0.062) p<0.001^a	β (SE)= -0.251 (0.064) p<0.001
Mean Diffusivity	Mean	β (SE)= -0.195 (0.109) p=0.077	β (SE)= -0.243 (0.064) p<0.001	β (SE)= -0.465 (0.066) p<0.001^a	β (SE)= -0.184 (0.015) p<0.001	B (SE)= -0.212 (0.051) p<0.001
	Histogram Width	β (SE)= -0.189 (0.135) p=0.166	β (SE)= -0.321 (0.077) p<0.001	β (SE)= -0.211 (0.065) p<0.001	β (SE)= -0.287 (0.066) p<0.001	β (SE)= -0.291 (0.063) p<0.001
Axial Diffusivity	Mean	β (SE)= -0.245 (0.125) p=0.053	β (SE)= -0.320 (0.072) p<0.001	β (SE)= -0.569 (0.065) p<0.001^a	β (SE)= -0.230 (0.047) p<0.001	β (SE)= -0.281 (0.058) p<0.001
	Histogram Width	β (SE)= -0.026 (0.136) p=0.852	β (SE)= -0.088 (0.057) p=0.125 ^a	β (SE)= -0.271 (0.066) p<0.001	β (SE)= -0.126 (0.067) p=0.060 ^a	β (SE)= -0.155 (0.068) p=0.023^a

Radial Diffusivity	Mean	β (SE)= -0.172 (0.104) p=0.101	β (SE)= -0.210 (0.061) p<0.001	β (SE)= -0.360 (0.053) p=0.014	β (SE)= -0.195 (0.051) p<0.001	β (SE)= -0.183 (0.049) p<0.001
	Histogram Width	β (SE)= -0.294 (0.131) p=0.027	β (SE)= -0.350 (0.075) p<0.001	β (SE)= -0.197 (0.061) p=0.001	β (SE)= -0.323 (0.062) p<0.001	β (SE)= -0.338 (0.061) p<0.001

results in bold are significant, ^aRobust regression; adjusting for GA at scan and sex

Table 5. Permutation testing of the relationship between GA at birth and dMRI metric means and histogram widths				
Diffusion metric	Measure	Effect of GA at birth in dHCP	Effect of GA at birth in harmonized data	T-test difference between regression coefficients
Fractional Anisotropy	Mean	0.098 (0.096-0.100) ±0.095	0.080 (0.078-0.081) ±0.088	t (19998)= 13.9 p<0.001
	β (95%CI) ± SD	Histogram Width -0.254 (-0.256-0.251) ±0.121	-0.254 (-0.257-0.252) ±0.113	t (19998)= 0.00 p>0.999
Mean Diffusivity	Mean	-0.261 (-0.263-0.259) ±0.105	-0.227 (-0.229-0.225) ±0.100	t (19998)= 23.4 p<0.001
	β (95%CI) ± SD	Histogram Width -0.341 (-0.344-0.339) ±0.136	-0.309 (-0.312-0.306) ±0.136	t (19998)= 16.6 p<0.001
Axial Diffusivity	Mean	-0.335 (-0.337-0.332) ±0.115	-0.287 (-0.289-0.285) ±0.112	t (19998)= 29.9 p<0.001
	β (95%CI) ± SD	Histogram Width -0.188 (-0.191-0.186) ±0.122	-0.148 (-0.151-0.146) ±0.123	t (19998)= 23.1 p<0.001
Radial Diffusivity	Mean	-0.229 (-0.231-0.227) ±0.101	-0.200 (-0.201-0.198) ±0.095	t (19998)= 20.9 p<0.001
	β (95%CI) ± SD	Histogram Width -0.352 (-0.354-0.349) ±0.134	-0.343 (-0.346-0.341) ±0.132	t (19998)= 4.78 p<0.001
Robust regression adjusting for gestational age at scan and sex; results in bold are significant				