



City Research Online

City, University of London Institutional Repository

Citation: Wade, R. G., Tam, W., Perumal, A., Pepple, S., Griffiths, T. T., Flather, R., Haroon, H. A., Shelley, D., Plein, S., Bourke, G. & et al (2023). Comparison of distortion correction preprocessing pipelines for DTI in the upper limb. *Magnetic Resonance in Medicine*, doi: 10.1002/mrm.29881

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/31620/>

Link to published version: <https://doi.org/10.1002/mrm.29881>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Comparison of distortion correction preprocessing pipelines for DTI in the upper limb

Ryckie G. Wade^{1,2}   | Winnie Tam¹  | Antonia Perumal¹  |
 Sophanit Pepple¹  | Timothy T. Griffiths^{1,2}  | Robert Flather^{1,2}  |
 Hamied A. Haroon³  | David Shelley⁴ | Sven Plein⁵  | Grainne Bourke^{1,2}  |
 Irvin Teh⁵ 

¹Leeds Institute for Medical Research, University of Leeds, Leeds, UK

²Department of Plastic, Reconstructive and Hand Surgery, Leeds Teaching Hospitals Trust, Leeds, UK

³Division of Psychology, Communication & Human Neuroscience, The University of Manchester, Manchester, UK

⁴Leeds Teaching Hospitals Trust, Leeds, UK

⁵Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK

Correspondence

Ryckie G. Wade, The Advanced Imaging Centre, Leeds General Infirmary, Leeds, LS1 3EX, UK.

Email: ryckiewade@nhs.net

Funding information

Arthritis Research UK, Grant/Award Number: ARUK-21078; British Heart Foundation, Grant/Award Numbers: BHF-SP/14/7/31351, CH/16/2/32089; Department of Health and Social Care; Medical Research Council, Grant/Award Number: MR/M008991/1; National Institute for Health and Care Research, Grant/Award Number: CL-2021-02-002; National Institute for Health Research, Grant/Award Number: DRF-2018-11-ST2-028; National Institute for Health Research Leeds Biomedical Research Centre; University of Leeds Advanced Imaging Centre; University of Leeds Medical School

Abstract

Purpose: DTI characterizes tissue microstructure and provides proxy measures of nerve health. Echo-planar imaging is a popular method of acquiring DTI but is susceptible to various artifacts (e.g., susceptibility, motion, and eddy currents), which may be ameliorated via preprocessing. There are many pipelines available but limited data comparing their performance, which provides the rationale for this study.

Methods: DTI was acquired from the upper limb of healthy volunteers at 3T in blip-up and blip-down directions. Data were independently corrected using (i) FSL's TOPUP & eddy, (ii) FSL's TOPUP, (iii) DSI Studio, and (iv) TORTOISE. DTI metrics were extracted from the median, radial, and ulnar nerves and compared (between pipelines) using mixed-effects linear regression. The geometric similarity of corrected $b = 0$ images and the slice matched T1-weighted (T1w) images were computed using the Sørensen-Dice coefficient.

Results: Without preprocessing, the similarity coefficient of the blip-up and blip-down datasets to the T1w was 0.80 and 0.79, respectively. Preprocessing improved the geometric similarity by 1% with no difference between pipelines. Compared to TOPUP & eddy, DSI Studio and TORTOISE generated 2% and 6% lower estimates of fractional anisotropy, and 6% and 13% higher estimates of radial diffusivity, respectively. Estimates of anisotropy from TOPUP & eddy versus TOPUP were not different but TOPUP reduced radial diffusivity by 3%. The agreement of DTI metrics between pipelines was poor.

Conclusions: Preprocessing DTI from the upper limb improves geometric similarity but the choice of the pipeline introduces clinically important variability in diffusion parameter estimates from peripheral nerves.

KEYWORDS

anisotropy, diffusion, diffusivity, distortion, DTI, magnetic resonance, median, nerve, pre-processing, radial, ulnar

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

1 | INTRODUCTION

Peripheral neuropathy and nerve injury are common, affecting 1 in 10 adults over the age of 40.¹ Diffusion-weighted MRI (dMRI) characterizes tissue microstructure and provides reproducible²⁻⁵ proxy measures of nerve health which are sensitive to axon type, diameter, density, myelination, and organization.⁶⁻⁹ DTI is the most prevalent form of dMRI and this is typically acquired by single-shot echo-planar imaging (EPI). However, EPI is susceptible to geometric and intensity distortions due to a combination of susceptibility-induced field inhomogeneities, eddy currents, subject motion, and low bandwidth in the phase-encoding direction.

Preprocessing is a multi-step process concerned with the correction of geometric and signal distortions which are prevalent in EPI data.¹⁰ The aim is to improve geometric fidelity and minimize false negatives, without increasing false positives in the postprocessing (analysis) phase. Although it is widely accepted that preprocessing should be performed because it improves the accuracy of dMRI metrics and tractography, there are no accepted standards. Consequently, practices and pipelines vary, which adversely affects the alignment of diffusion-weighted and anatomical images,¹¹ generating important differences in tensor based metrics¹² and tractograms.² Collectively, this negatively impacts the reproducibility of dMRI studies.^{13,14}

The majority of the dMRI community agree that preprocessing pipelines should be standardized.¹⁵ Furthermore, over 80% believe that corrections for artifacts arising from subject motion, eddy currents, field inhomogeneities, and thereafter, *b*-matrix rotation are required.¹⁵ Several studies have shown that phase-encoding-based methods yield the best corrections of the above artifacts.^{14,16,17} Ideally, full dMRI datasets are acquired in opposing phase-encoding (PE) directions (that is, all images in both blip-up and blip-down directions)¹² although similar results can be achieved by acquiring non-diffusion-weighted (DW) data with reversed PE or via deep-learning.¹⁸ There are several open-source software packages that perform corrections for distortions arising from some or all of the following: motion, eddy currents, and susceptibility-induced artifacts, using data with opposing phase encoding, such as *animaDistortionCorrection*,¹⁹ *animaBMDistortionCorrection*,²⁰ *DR-BUDDI* from the TORTOISE suite,²¹⁻²³ *DSI Studio*,²⁴ *EPIC*,²⁵ *HySCO*,²⁶ and *TOPUP*²⁷ and *eddy*²⁸ from the FMRIB Software Library (FSL). However, there is uncertainty about which pipeline offers the best performance and most packages were developed for brain imaging, so there is limited work on non-brain real-world data.²⁹

In this study, we compared the performance of the three most common software packages that offer

phase-encoding-based preprocessing of DTI, namely FSL's suite (which is also embedded into the most popular postprocessing package worldwide, MRtrix3³⁰), DSI Studio, and TORTOISE.

2 | METHODS

This cross-sectional study was designed and reported in accordance with the STROBE and STARD guidance, taking into account the domains of the QUADAS-2³¹ and PRISMA-DTA³² tools. Approval was provided by the National Health Research Authority (ID 19/NW/0324), and written informed consent was obtained from all participants.

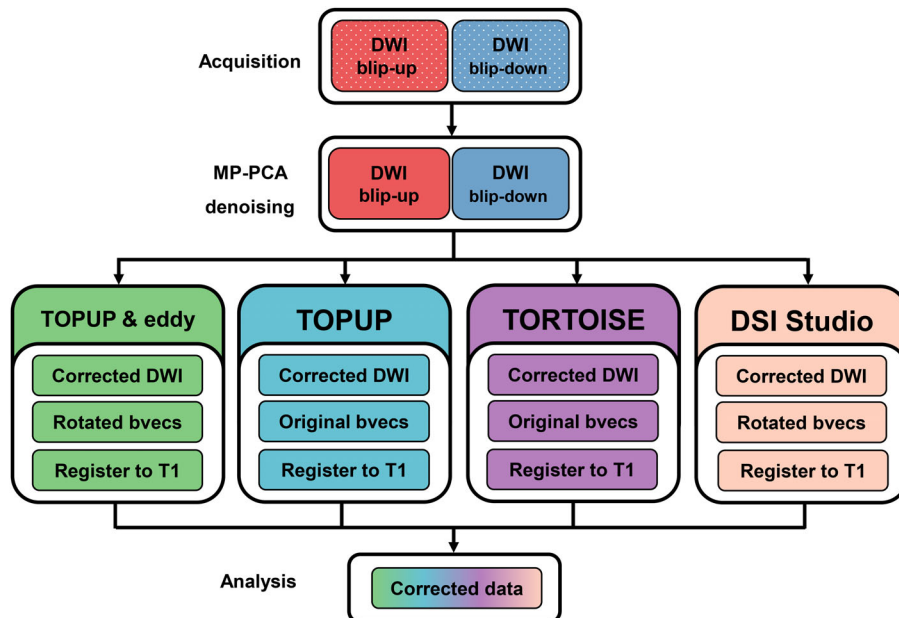
2.1 | Recruitment

We included 13 healthy participants (8 males and 5 females) of mean age 29 y (SD 6.22, range 21–40). Their mean height was 174 cm (SD 9) and weight 75 kg (SD 17), giving a mean body mass index of 25 (SD 5). Volunteers were recruited opportunistically between July 2019 and March 2020. We excluded volunteers who had non-MR safe implants, peripheral neuropathy, metallic implants near the elbow or claustrophobia.

2.2 | Image acquisition

DTI data were acquired at a field strength of 3T using a MAGNETOM Prisma (Siemens Healthcare, Erlangen, Germany) and a single-shot EPI sequence. Participants were scanned prone, with the shoulder flexed and elbow extended. The elbow was positioned as close to the isocenter of the magnet as comfortably possible. A four-channel flexible coil was wrapped around the elbow and secured with strapping. Fifty-five contiguous axial slices of 3 mm thickness were acquired, at an in-plane resolution of 1.5 mm².³³ The FOV was reduced to 192 × 78 mm (frequency × phase) using ZOOMit (TimTX TrueShape) with TrueForm B1 shim. We applied 30 non-collinear monopolar diffusion encoding gradients using a Jones scheme³⁴ with the following parameters: *b*-value 800 s/mm² ($\Delta = 19.0$ ms, $\delta = 13.2$ ms), four interleaved non-DW (*b* = 0) images, TE 74 ms, TR 7800 ms, effective echo spacing 0.97 ms and EPI (“turbo”) factor 52 (giving an echo train length of ~50 ms), GRAPPA off, 6/8 partial Fourier, receiver bandwidth 1184 Hz, distortion correction off and strong fat saturation. Four signals/averages (totaling 120 DWIs and 16 interleaved *b* = 0 images) were acquired which required 17 min 50 s. The sequence

FIGURE 1 Steps for comparing preprocessing pipelines. Note that both TOPUP & eddy are part of the FMRIB Software Library (FSL). By default, both FSL's eddy and DSI Studio automatically rotated the b -matrix³⁷ whereas the analogous function in TORTOISE did not. TOPUP alone did not alter the b -matrix. When registering the dMRI data to the space of T1, DSI Studio automatically further transformed to the b -matrix and saved the newly rotated b -matrix within the native .fib file for analysis.



was repeated with the (right–left) PE direction reversed. This was supplemented by slice-matched T1-weighted (T1w) images.

2.3 | Preprocessing

DICOMs were converted to nifti using `dcm2nii`³⁵ and denoised by MP-PCA.³⁶ Thereafter, data were passed to each pipeline for concatenation and correction (Figure 1).

In DSI Studio²⁴ (using the November 16th 2021 release), we used the “Correct AP-PA scans” option whereby the correction is based on the computed cumulated spin density along the phase encoding direction and applied using point-to-point mapping to restore the unwrapped distribution. Next, we applied the “motion correction” option which applies a linear registration based on mutual information between the $b=0$ images and DWIs. The summative transformation is then applied to the b -matrix. The corrected data were saved in the software specific.fib format (which includes the corrected b -matrix within) and also exported in nifti format.

In FSL v6.0.2, the $b=0$ images were extracted and passed to TOPUP, using the default configuration file (b02b0.cnf). To test the performance of susceptibility correction only within FSL, the `applyTOPUP` command was used, supplying the blip-up and blip-down data and default parameters.

Again, in FSL v6.0.2, after running TOPUP with the default configuration file (b02b0.cnf), a binary mask (removing only noise) was generated from the corrected mean $b=0$ images, and these outputs were passed to eddy (`eddy_cuda`) with the following

options and configurations. The eddy current-induced fields were modeled using a quadratic function (`--flm=quadratic`). Data of opposing phase were concatenated by least-squares resampling (`--resamp=lsr`), which is the default. Outlier slices (with signal intensities at least four SDs from the expected) were replaced with predictions made by Gaussian Processes using the `--repol` option³⁸; according to the outlier reports, a median 1 slice per volunteer (0.3% of the total slices per volunteer, range 0–11 slices) was replaced. We also enabled the slice-to-volume motion correction using the `--mporder` option³⁹ with 15 degrees of freedom and `--estimate` `move_by_susceptibility`⁴⁰ whereby susceptibility-induced field change due to subject motion is estimated and corrected using a first-order Taylor expansion of the static field, with respect to pitch and roll.

Within TORTOISE, DIFFPREP was applied, and data were combined and corrected using Diffeomorphic Registration for Blip-Up blip-Down Diffusion Imaging (DR_BUDDI)²² using the default settings. The mean $b=0$ images were selected as the reference (T2w) dataset. In two volunteers, some slices at the ends of the stack (in the proximal arm and distal forearm regions) could not be corrected, so these specific slices were excluded from postprocessing. We could not identify a rotated b -table or transformation matrix in the output folder of DR_BUDDI, so the original b -vectors were used.

2.4 | Postprocessing

The corrected datasets from each pipeline were imported to DSI Studio. Datasets were automatically registered to

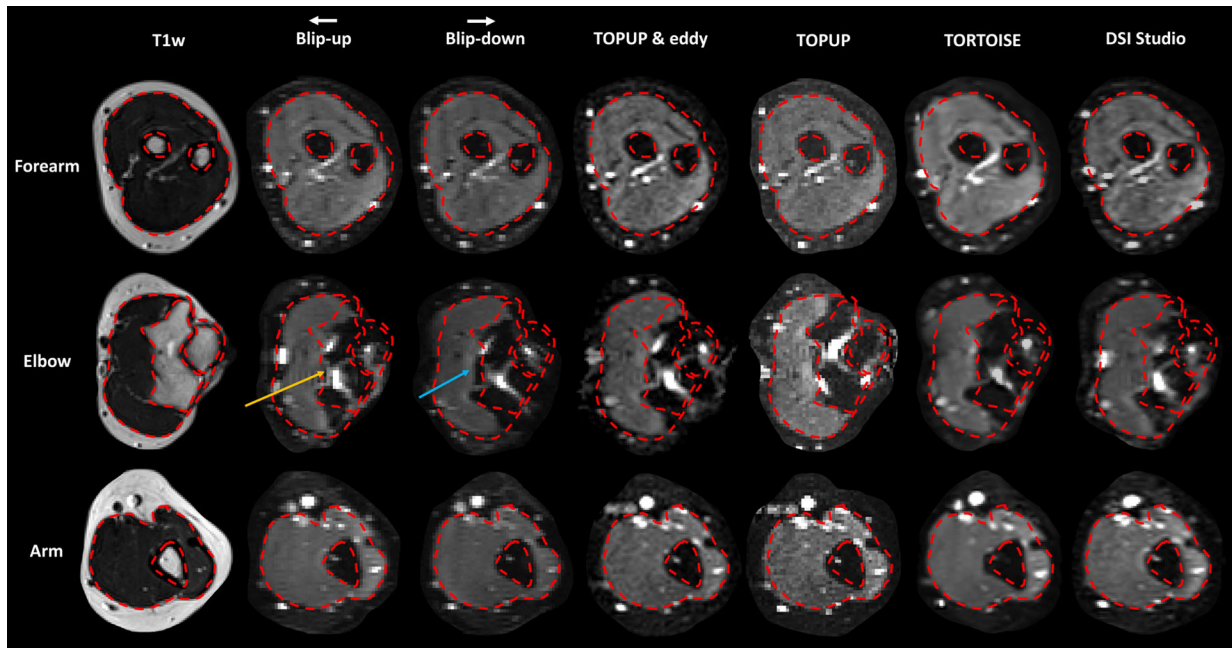


FIGURE 2 A montage showing slice matched T1-weighted images, $b=0$ images from the blip-up and blip-down dMRI acquisitions, and corrected $b=0$ images from each pipeline, in the right limb of a 30-y-old female. The phase-encoding direction (white arrow) and consequential areas of signal compression/pile-up (gold arrow) and dilation (blue arrow) are shown. The red dotted lines were drawn on the T1w image and overlaid on the $b=0$ images.

the T1w data using a rigid body transformation and the same transformation was automatically applied to the b -matrix. Diffusion was quantified using restricted diffusion imaging⁴¹ and reconstructed using Generalized Q-Space Imaging (GQI).⁴² This model-free approach was selected to compare the four preprocessing approaches because peripheral nerves are more conspicuous on the resultant quantitative anisotropy (QA) maps, as compared to traditional tensor-based maps. Also, GQI is readily applicable to numerous different diffusion sampling schemes, the outputs are comparable to more complex q-space methods and it generates a spin-density function that is the closest to reality.⁴²

2.5 | Segmentation of the limb

Given the variability of coil position and therefore proximal/distal coverage, we classified a 5.7 cm section (19 axial slices) centered on the radiohumeral joint as the “Elbow” region. Measurements proximal to this were classified as within the “Arm” and measurements distal were classified as the “Forearm”.

2.6 | Measuring the similarity of the T1w and corrected DTI images

To quantify the morphological similarity of the corrected DTI datasets to the anatomical (T1w) reference image,

regions of interest (ROIs) were drawn to enclose the sub-fascial structures of the upper limb (Figures 2 and 3). The regions were drawn manually on the $b=0$ images of the raw blip-up, blip-down, and corrected dMRI datasets, as well as the slice-matched T1w images, by RGW. This was performed on matching slices in the arm, elbow, and forearm. The volumetric overlap of these regions were binarized and then used to calculate the Sørensen-Dice similarity coefficient.^{43,44}

2.7 | Extraction of metrics

To extract metrics from the median, radial, and ulnar nerves within each of the corrected datasets, ROIs were drawn over the center of the ulnar, median, and radial nerves on every QA map by three researchers (WT, AP, and SP; Figure S1). All ROIs were checked independently by RGW, TG and RF. Around the elbow, the normal cross-sectional area of the ulnar nerve is $\sim 7\text{mm}^2$,^{45,46} the median $\sim 7\text{mm}^2$, and the radial nerve is $\sim 5.1\text{mm}^2$.⁴⁶ To minimize partial volume effects, the ROI was limited to 1.5mm^2 (one voxel) and centered over the cross-section of the nerve, which typically had the highest regional QA value. The following metrics were then extracted from each ROI throughout the length of the nerve: fractional anisotropy (FA), radial diffusivity (RD), axial diffusivity (AD), and mean diffusivity (MD) and their corresponding maps are shown in Figure S2.

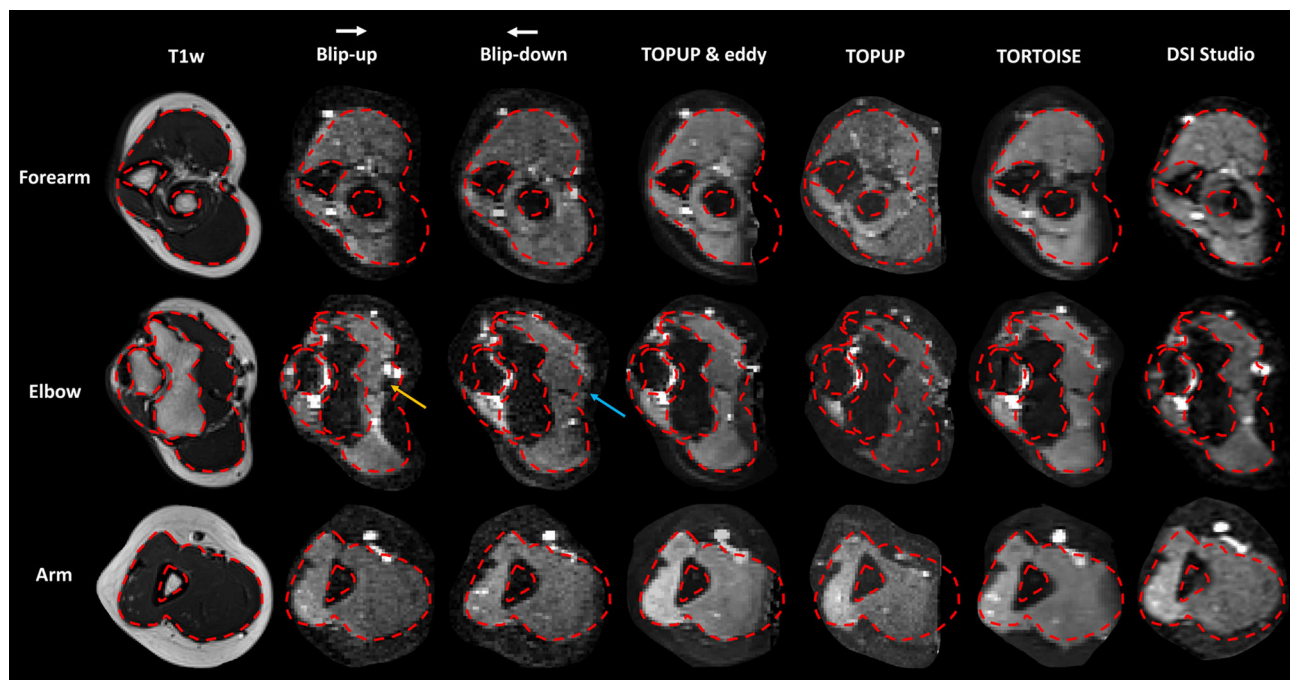


FIGURE 3 A montage showing slice matched T1-weighted images, $b = 0$ images from the blip-up and blip-down dMRI acquisitions, and corrected $b = 0$ images from each pipeline, in the left limb of a 39-y-old female. The phase-encoding direction (white arrow) and consequential areas of signal compression/pile-up (gold arrow) and dilation (blue arrow) are shown. The red dotted lines were drawn on the T1w image and overlaid on the $b = 0$ images.

2.8 | Analysis

The raw data are available open source at <https://osf.io/z29m6/> (last accessed September 7, 2023). Data were analyzed using Stata v16/MP (StataCorp LLC, Texas). Normality was confirmed by visualization of the distribution of data, so scaled variables are represented by the mean (and SD). The Sørensen-Dice coefficients were skewed in one subgroup, so are represented globally by the geometric mean and 95% confidence interval (CI). The effect of different preprocessing pipelines on the Sørensen-Dice coefficient was explored using mixed-effects linear regression, given that the residuals were normally distributed. The dependent variables in four separate models were the DTI metrics (FA, MD, RD, and AD). The fixed effect in each model was the method of processing (blip-up, blip-down, TOPUP & eddy, TOPUP, DSI Studio, or TORTOISE). Restricted maximum likelihood was used to estimate the cluster-level (volunteer) variance. The variance and covariance parameters were unstructured and, so, distinctly estimated. The outputs of these models are shown as cubic splines, formatted using grstyle.^{47,48} To summarize agreement between pipelines, the intraclass correlation and Pearson's r coefficients were computed and shown alongside Raincloud⁴⁹ and Bland-Altman plots. In line with calls for the abolition of p -values, we have minimized their use and avoided the term “statistical

significance”,^{50,51} instead focussing on how the findings might translate to future real-world imaging scenarios.

3 | RESULTS

Without preprocessing, the geometric mean Sørensen-Dice coefficient of the blip-up and blip-down datasets to the T1w was 0.80 (CI 0.78–0.83) and 0.79 (CI 0.75–0.83), respectively. Preprocessing improved the similarity by a mean of 1% (mean Sørensen-Dice of TOPUP & eddy 0.81 [CI 0.77–0.83], TOPUP alone 0.83 [CI 0.81–0.86], DSI Studio 0.80 [CI 0.77–0.83], and TORTOISE 0.81 [CI 0.78–0.85]; Figure 4) with no meaningful differences between pipelines.

Based on the Sørensen-Dice coefficients, distortions were worst in the elbow region (0.74 [CI 0.71–0.77]) but similar in the arm (0.82 [CI 0.79–0.86]) and forearm (0.83 [CI 0.76–0.90], Figure S3). Overall, distortions were better corrected in the arm and forearm (Sørensen-Dice coefficients were 8% and 9% higher, respectively) than in the elbow region, but these differences were not clinically meaningful (Figure S4).

As shown in Table S1 and Figure 5, DTI metrics from the median, radial, and ulnar nerves differed substantially according to the preprocessing pipeline used. In comparison to TOPUP & eddy, DSI Studio

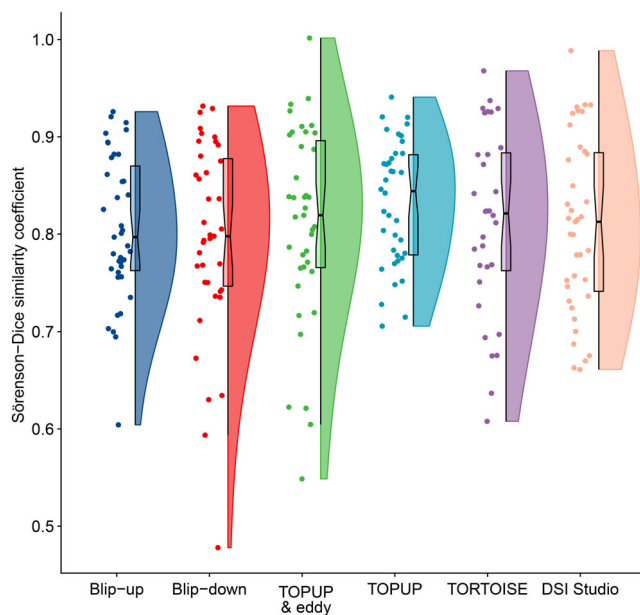


FIGURE 4 Raincloud plot showing the Sørensen-Dice similarity coefficient of the raw blip-up and blip-down and preprocessed datasets against the T1w reference images.

and TORTOISE produced estimates of the FA which were 2% (CI 2–3) and 6% (CI 5–7) lower, respectively (Figure 5). There was no differences in FA between TOPUP & eddy and TOPUP alone. The agreement in FA between pipelines was generally poor (Figures S5–S10, Pearson's $r < 0.6$; ICC < 0.2). Compared to TOPUP & eddy, TOPUP generated lower estimates of MD ($0.075 \times 10^{-3} \text{ mm}^2/\text{s}$ [CI 0.053×10^{-3} – 0.096×10^{-3}]), whereas DSI Studio yielded higher estimates of MD ($0.032 \times 10^{-3} \text{ mm}^2/\text{s}$ [CI 0.012×10^{-3} – 0.052×10^{-3}]). MD was not different between TOPUP & eddy and TORTOISE (Figure 5). In comparison to TOPUP & eddy, DSI Studio and TORTOISE both produced higher RD values ($0.053 \times 10^{-3} \text{ mm}^2/\text{s}$ [CI 0.033×10^{-3} – 0.074×10^{-3}] and 0.136 [CI 0.113×10^{-3} – 0.159×10^{-3}]), respectively, whereas TOPUP reduced the RD by 0.041 (CI 0.018×10^{-3} – 0.064×10^{-3}) (Figure 5). Compared to TOPUP & eddy, estimates of AD were lowered in TOPUP, DSI Studio, and TORTOISE by $0.145 \times 10^{-3} \text{ mm}^2/\text{s}$ (CI 0.116 – 0.173), $0.010 \times 10^{-3} \text{ mm}^2/\text{s}$ (CI 0.016×10^{-3} – $0.037 \times 10^{-3} \text{ mm}^2/\text{s}$), and $0.234 \times 10^{-3} \text{ mm}^2/\text{s}$ (CI 0.205×10^{-3} – $0.263 \times 10^{-3} \text{ mm}^2/\text{s}$), respectively (Figure 5).

4 | DISCUSSION

This study suggests that software used for preprocessing DTI improves the geometric accuracy of dMRI data from upper limb with respect to the anatomical ground truth, but this comes at the expense of introducing large

differences in the diffusion metrics extracted from peripheral nerves. This is clinically important because the magnitude of variation in DTI parameters between preprocessing software is of the magnitude that might be attributable to disease or injury.^{52,53} Consequently, in the absence of standardization of methods used by pipelines, readers should be aware of the potential for preprocessing to bias the output.

It is widely accepted that preprocessing of dMRI data derived from EPI is beneficial¹⁷ and improves the re-test reliability of group studies.⁵⁴ Furthermore, EPI artifacts are prevalent throughout the magnitude image (not just localized to the sources of classically cited distortion, e.g., air-tissue interfaces); hence, preprocessing benefits datapoints throughout the entire volume.⁵⁵ However, there are numerous software packages available for preprocessing, each with different approaches and user-specified options that can be deployable in any order or combination. This freedom means that the plausible combinations is substantial and arguably at least 15 factorial.¹⁰ It is well known that variability in preprocessing generates differences in the alignment of dMRI datasets with anatomical images,¹¹ tensor-based metrics,¹² and tractograms² in the brain. This in turn reduces the external validity of studies.^{13,14} Although recent works suggest that there are no meaningful differences in the outputs of different pipelines,⁵⁶ controversy still exists. Consequently, the majority (55%) of the dMRI community feels that preprocessing pipelines should be standardized.¹⁵ Of those advocating for standardization, 80% stated that the minimum should include corrections for artifacts arising from subject motion, eddy currents, field inhomogeneities, and thereafter, *b*-matrix rotation should be performed.¹⁵ Given the importance and complexity of this topic, the ISMRM Diffusion Study Group assembled a research collaborative including 232 international scientists and clinicians. The aim was to survey the community regarding typical practices, and thereafter, collaborators were invited to preprocess 13 multi-shell spin-echo EPI dMRI datasets from the brain, from multi-vendor 3T systems (GE SIGNA, Siemens Connectom and Philips Achieva) from multiple sites and sessions. The outputs from this collaborative study should provide important insight into the scale of the problem¹⁵ and lay the foundations for how preprocessing might be standardized.⁵⁷

Importantly, we used ZOOMit (Siemens reduced FOV product), which deploys a dynamic excitation pulse to reduce the length of the EPI train. This has two major advantages, namely shorter TEs and distortion reduction. This may explain why we observed that preprocessing only improved the Sørensen-Dice similarity coefficient of the DTI and T1w reference by a mean 1%. It is expected that other vendors small FOV products (e.g., GE's FOCUS or

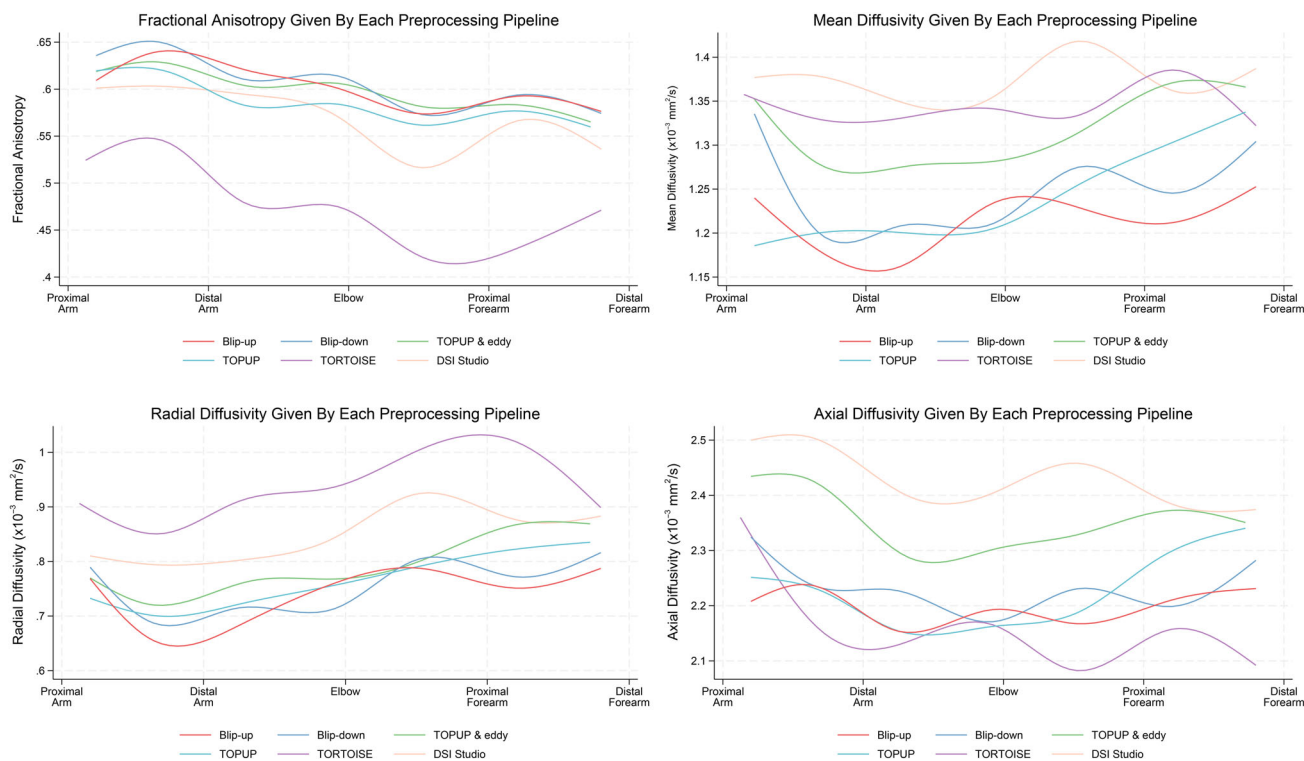


FIGURE 5 Fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity (plotted as cubic splines) in the peripheral nerves of the upper limb, according to their location in the limb and what (if any) pipeline was used for preprocessing.

Philip's iZOOM) may perform differently and moreover, non-reduced-FoV EPI data may have more severe distortions and therefore, preprocessing may be of greater value in correcting the geometry. We recommend that future studies that compare the performance of preprocessing pipelines in extremity dMRI include data acquired using non-reduced FOV products and across the entire volume.

Although one pipeline (or more specifically, a given combination of corrections within a given software pipeline) may statistically outperform other pipelines (e.g., better similarity coefficients or better agreement with reference phantoms), there are other factors that users must consider when selecting a software pipeline. These might include the cost of the software, user interface, stability, compatibility with different operating systems, exploitation of hardware acceleration, developer support and integration with other popular tools. In relation to the tools compared in this study, there are some strengths and weaknesses as we see it. DSI Studio has a user-friendly graphical interface that requires no knowledge of coding (e.g., Python, Unix, etc), the installation is easy, and the software is stable across Windows, MacOS, and various distros of Linux. In contrast, both FSL and TORTOISE warrant some knowledge of terminal-based scripting to install and use, which may prohibit some users. Additional difficulties are presented to users of FSL on Microsoft Windows because it requires a virtual machine or the Windows Subsystem

for Linux (WSL) and allied graphical software to run. Furthermore, the graphics processing unit (GPU)-accelerated features of FSL require additional libraries and drivers, which may be challenging for users of MacOS given the recent removal of NVIDIA compatibility. It is commendable that both DSI Studio and FSL offer timely and detailed support for users via their respective forums and training courses, which have run for several years. However, user support for TORTOISE appears to have been discontinued and no training course has been advertised for several years. Beyond the empirical performance of each software, there are wider issues that prospective users should consider before committing to a particular software package.

4.1 | Limitations

As there is no reference standard for DTI parameter estimation, we are limited to relative comparisons, whereas we would ideally compute the absolute biases/errors from the truth. Phantoms could potentially help provide reference measurements by which to evaluate the performance of the different pipelines but mimicking the microstructure and spatial context of peripheral nerves is a challenging task that requires further development. Equally, fixed post-mortem specimens may be helpful, but they too present unique issues, for example, reduced

diffusivity and T_2 , which renders clinical EPI sequences inadequate; the difficulty maintaining physiological body temperature; the absence of artifacts related to flow, movement and breathing, which would be present in in-vivo imaging; the degradation of microstructure in the period between death and fixation; and potential tissue changes associated with choice of fixation protocol. Overall, we acknowledge that the absence of a ground truth diminishes the impact of this study. We used the November 16th 2021 release of DSI Studio²⁴ for this study; readers should note that this software suite has a high version turning rate, and since our work was completed, the “Chen” version was released, which includes the option to use FSL’s TOPUP²⁷ and eddy²⁸ regimes for correcting datasets with reversed $b=0$ images. Moreover, this function has been optimized to exploit multiple central processing unit (CPU) cores, it now automatically detects and configures the phase-encoding direction (which users otherwise must prescribe manually in FSL), masks/segments of the data, and deploys applytopup if eddy cannot be run. All these functions are embedded within DSI Studio, meaning that no additional downloads/installations are needed. Moreover, this is now all achievable in Windows (without the requirement for virtualization/WSL or a native installation of FSL) and Mac (with a native installation of FSL), with the full version of eddy (to concatenated full dMRI datasets of opposing phase) forecast to be incorporated into DSI Studio soon. We enabled the replacement of outlier slices in the FSL pipeline, which is the only inter-pipeline difference of note; although only a median of one slice per volunteer was replaced (representing 0.3% of the total slices per volunteer), this may have had a minor impact on the resultant DTI metrics. We expect that our findings may be influenced by the placement of ROIs, given that metrics were extracted from one voxel overlying each nerve, per slice. We tried to guard against this by having ROIs independently checked by multiple researchers and amended by consensus. Also, we averaged data across numerous slices, which should mitigate against such noise/variability. Nonetheless, it is plausible that variation in ROI placement explains some of the observed differences. Importantly, most packages were developed for brain imaging, principally to prepare and correct voxels in a single masked brain as part of a system or network of neurons, rather than multiple structures in the body contained in the same image where the respective signals should be separate. Therefore, there is limited work on non-brain real-world data in this context. Although many authors have written extensively on the importance of using a rotated b -matrix,^{37,58} we could not find a rotated b -matrix or transformation matrix in the outputs of DR_BUDDI and so we used the raw b -vectors, which may explain some differences. Finally, the smoothness of data from each pipeline

appeared to be different, whereby TORTOISE data were noticeably smoother than others (although its documentation does not describe a smoothing function), and this may translate to differences in DTI metrics.

5 | CONCLUSIONS

Preprocessing dMRI data from the upper limb improves geometric accuracy but the choice of the pipeline appears to introduce a source of clinically important variability in the diffusion parameter estimates from peripheral nerves.

FUNDING INFORMATION

Ryckie Wade is an Academic Clinical Lecturer funded by the National Institute for Health Research (NIHR, CL-2021-02-002). This research is also supported by the NIHR Leeds Biomedical Research Centre and University of Leeds Advanced Imaging Centre, which is funded by the Medical Research Council (MR/M008991/1) with support from the British Heart Foundation (BHF-SP/14/7/31351) and Arthritis Research UK (ARUK-21078). The views expressed are those of the author(s) and not necessarily those of the United Kingdom’s National Health Service, NIHR or Department of Health. Sven Plein is funded by a British Heart Foundation Chair (CH/16/2/32089). Timothy Griffiths was supported by the University of Leeds Medical School Alumni Footsteps Scholarship.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The raw data are available open source at <https://osf.io/z29m6>.

ORCID

Ryckie G. Wade  <https://orcid.org/0000-0001-8365-6547>

Winnie Tam  <https://orcid.org/0000-0002-6241-4810>

Antonia Perumal  <https://orcid.org/0000-0001-5550-6349>

Sophanit Pepple  <https://orcid.org/0000-0002-2782-3192>

Timothy T. Griffiths  <https://orcid.org/0000-0002-5993-8525>

Robert Flather  <https://orcid.org/0000-0003-0458-1802>

Hamied A. Haroon  <https://orcid.org/0000-0002-3506-8430>

Sven Plein  <https://orcid.org/0000-0002-0997-4384>

Grainne Bourke  <https://orcid.org/0000-0003-0059-8543>

Irvin Teh  <https://orcid.org/0000-0002-6705-3129>

TWITTER

Ryckie G. Wade  [ryckiewade](https://twitter.com/ryckiewade)

REFERENCES

- Hicks CW, Wang D, Windham BG, Matsushita K, Selvin E. Prevalence of peripheral neuropathy defined by monofilament insensitivity in middle-aged and older adults in two US cohorts. *Sci Rep*. 2021;11:19159.
- Nath V, Schilling KG, Parvathaneni P, et al. Tractography reproducibility challenge with empirical data (TraCED): the 2017 ISMRM diffusion study group challenge. *J Magn Reson Imaging*. 2020;51:234-249.
- Vavasour IM, Meyers SM, Mädler B, et al. Multicenter measurements of T_1 relaxation and diffusion tensor imaging: intra and intersite reproducibility. *J Neuroimaging*. 2019;29:42-51.
- Prohl AK, Scherrer B, Tomas-Fernandez X, et al. Reproducibility of structural and diffusion tensor imaging in the TACERN multi-center study. *Front Integr Neurosci*. 2019;13:24.
- Kimura M, Yabuuchi H, Matsumoto R, et al. The reproducibility of measurements using a standardization phantom for the evaluation of fractional anisotropy (FA) derived from diffusion tensor imaging (DTI). *Magn Reson Mater Phys Biol Med*. 2019;33:15-19.
- Andersson G, Orädd G, Sultan F, Novikov LN. In vivo diffusion tensor imaging, diffusion kurtosis imaging, and tractography of a sciatic nerve injury model in rat at 9.4T. *Sci Rep*. 2018;8:12911.
- Friedrich P, Fraenz C, Schlüter C, et al. The relationship between axon density, myelination, and fractional anisotropy in the human corpus callosum. *Cereb Cortex*. 2020;30:2042-2056.
- Heckel A, Weiler M, Xia A, et al. Peripheral nerve diffusion tensor imaging: assessment of axon and myelin sheath integrity. *PLoS One*. 2015;10:e0130833.
- Pušnik L, Serša I, Umek N, Cvetko E, Snoj Ž. Correlation between diffusion tensor indices and fascicular morphometric parameters of peripheral nerve. *Front Physiol*. 2023;14:1070227.
- Tax CMW, Bastiani M, Veraart J, Garyfallidis E, Okan Irfanoglu M. What's new and what's next in diffusion MRI preprocessing. *Neuroimage*. 2022;249:118830.
- Irfanoglu MO, Walker L, Sarlls J, Marengo S, Pierpaoli C. Effects of image distortions originating from susceptibility variations and concomitant fields on diffusion MRI tractography results. *Neuroimage*. 2012;61:275-288.
- Brun L, Pron A, Sein J, Deruelle C, Coulon O. Diffusion MRI: assessment of the impact of acquisition and preprocessing methods using the BrainVISA-diffuse toolbox. *Front Neurosci*. 2019;13:536.
- Carp J. On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Front Neurosci*. 2012;6:149.
- Esteban O, Daducci A, Caruyer E, et al. Simulation-based evaluation of susceptibility distortion correction methods in diffusion MRI for connectivity analysis. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, Beijing, China; 2014: 738-741. doi:10.1109/ISBI.2014.6867976
- Veraart J. A Data-Driven Variability Assessment of Brain Diffusion MRI Preprocessing Pipelines. Proceedings of the Joint Annual ISMRM-ESMRMB Meeting 2022, London. Abstract 4983.
- Graham MS, Drobnjak I, Jenkinson M, Zhang H. *Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI*. Vol 12. PLoS One; 2017:e0185647.
- Gu X, Eklund A. Evaluation of six phase encoding based susceptibility distortion correction methods for diffusion MRI. *Front Neuroinform*. 2019;13:76.
- Schilling KG, Blaber J, Hansen C, et al. Distortion correction of diffusion weighted MRI without reverse phase-encoding scans or field-maps. *PLoS One*. 2020;15:e0236418.
- Voss HU, Watts R, Uluğ AM, Ballon D. Fiber tracking in the cervical spine and inferior brain regions with reversed gradient diffusion tensor imaging. *Magn Reson Imaging*. 2006;24:231-239.
- Hedouin R, Commowick O, Bannier E, et al. Block-matching distortion correction of echo-planar images with opposite phase encoding directions. *IEEE Trans Med Imaging*. 2017;36:1106-1115.
- Irfanoglu MO, Sarlls J, Nayak A, Pierpaoli C. Evaluating corrections for Eddy-currents and other EPI distortions in diffusion MRI: methodology and a dataset for benchmarking. *Magn Reson Med*. 2018;81:2774-2787.
- Irfanoglu MO, Modi P, Nayak A, Hutchinson EB, Sarlls J, Pierpaoli C. DR-BUDDI (diffeomorphic registration for blip-up blip-down diffusion imaging) method for correcting echo planar imaging distortions. *Neuroimage*. 2015;106:284-299.
- Pierpaoli C. TORTOISE: an integrated software package for processing of diffusion MRI data. *SMRM 18th Annu. Meet. Stockh. Swed. Abstr*. 1597 2009.
- Yeh F. DSI studio. Zenodo; 2021. doi:10.5281/zenodo.4978980
- Holland D, Kuperman JM, Dale AM. Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo planar imaging. *Neuroimage*. 2010;50:175-183.
- Ruthotto L, Mohammadi S, Heck C, Modersitzki J, Weiskopf N. Hyperelastic susceptibility artifact correction of DTI in SPM. In: Meinzer HP, Deserno T, Handels H, Tolxdorff T, eds. *Bildverarbeitung für die Medizin 2013*. Informatik aktuell, Springer. doi:10.1007/978-3-642-36480-8_60
- Andersson J, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*. 2003;20:870-888.
- Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*. 2016;125:1063-1078.
- Hancu I, Lee S-K, Hulsey K, et al. Distortion correction in diffusion-weighted imaging of the breast: performance assessment of prospective, retrospective, and combined (prospective + retrospective) approaches: distortion correction in diffusion-weighted imaging of the breast. *Magn Reson Med*. 2017;78:247-253.
- Tournier J-D, Smith R, Raffelt D, et al. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*. 2019;202:116137.
- Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
- McGrath TA, Alabousi M, Skidmore B, et al. Recommendations for reporting of systematic reviews and meta-analyses of diagnostic test accuracy: a systematic review. *Syst Rev*. 2017;6: 194.
- Chang K-V, Wu W-T, Han D-S, Özçakar L. Ulnar nerve cross-sectional area for the diagnosis of cubital tunnel syndrome: a meta-analysis of Ultrasonographic measurements. *Arch Phys Med Rehabil*. 2018;99:743-757.

34. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn Reson Med.* 1999;42:515-525.
35. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J Neurosci Methods.* 2016;264:47-56.
36. Veraart J, Fieremans E, Novikov DS. Diffusion MRI noise mapping using random matrix theory. *Magn Reson Med.* 2016;76:1582-1593.
37. Leemans A, Jones DK. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med.* 2009;61:1336-1349.
38. Andersson JLR, Graham MS, Zsoldos E, Sotiropoulos SN. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage.* 2016;141:556-572.
39. Andersson JLR, Graham MS, Drobnyak I, Zhang H, Filippini N, Bastiani M. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: within volume movement. *Neuroimage.* 2017;152:450-466.
40. Andersson JLR, Graham MS, Drobnyak I, Zhang H, Campbell J. Susceptibility-induced distortion that varies due to motion: correction in diffusion MR without acquiring additional data. *Neuroimage.* 2018;171:277-295.
41. Yeh F-C, Liu L, Hitchens TK, Wu YL. Mapping immune cell infiltration using restricted diffusion MRI. *Magn Reson Med.* 2017;77:603-612.
42. Yeh F-C, Wedeen VJ, Tseng W-YI. Generalized q-sampling imaging. *IEEE Trans Med Imaging.* 2010;29:1626-1635.
43. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *K Dan Vidensk Selsk.* 1948;5:1-34.
44. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945;26:297-302.
45. Boers N, Martin E, Mazur M, et al. Sonographic normal values for the cross-sectional area of the ulnar nerve: a systematic review and meta-analysis. *J Ultrasound.* 2023;26:81-88. doi:10.1007/s40477-022-00661-8
46. Fisse AL, Katsanos AH, Gold R, Pitarokoili K, Krogias C. Cross-sectional area reference values for peripheral nerve ultrasound in adults: a systematic review and meta-analysis—part I: upper extremity nerves. *Eur J Neurol.* 2021;28:1684-1691.
47. Jann B. Customizing stata graphs made easy (part 2). *Stata J.* 2018;18:786-802.
48. Jann B. Customizing stata graphs made easy (part 1). *Stata J.* 2018;18:491-502.
49. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res.* 2019;4:63.
50. Wasserstein RL, Lazar NA. The ASA Statement on p-values: context, process, and purpose. *Am Stat.* 2016;70:129-133.
51. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature.* 2019;567:305-307.
52. Griffiths TT, Flather R, Teh I, et al. Diffusion tensor imaging in cubital tunnel syndrome. *Sci Rep.* 2021;11:14982.
53. Rojoa D, Raheman F, Rassam J, Wade RG. Diffusion tensor imaging of the median nerve: a systematic review and meta-analysis of normal values in asymptomatic adults and how they change in carpal tunnel syndrome. *SSRN Electron J.* 2021;11:1-11.
54. Wang S, Peterson DJ, Gatenby JC, Li W, Grabowski TJ, Madhyastha TM. Evaluation of field map and nonlinear registration methods for correction of susceptibility artifacts in diffusion MRI. *Front Neuroinform.* 2017;11:17. doi:10.3389/fninf.2017.00017/full
55. Begnoche JP, Schilling KG, Boyd BD, Cai LY, Taylor WD, Landman BA. EPI susceptibility correction introduces significant differences far from local areas of high distortion. *Magn Reson Imaging.* 2022;92:1-9.
56. Kornaropoulos EN, Winzeck S, Rumetshofer T, et al. Sensitivity of diffusion MRI to white matter pathology: influence of diffusion protocol, magnetic field strength, and processing pipeline in systemic lupus erythematosus. *Front Neurol.* 2022;13:837385.
57. Schilling KG, Rheault F, Petit L, et al. Tractography dissection variability: what happens when 42 groups dissect 14 white matter bundles on the same dataset? *Neuroimage.* 2021;243:118502.
58. Pierpaoli C, Walker L, Irfanoglu MO, et al. TORTOISE: an integrated software package for processing of diffusion MRI data. Proceedings of Joint Annual ISMRM-ESMRMB Meeting. 2010. Abstract 1597.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. Examples of the regions of interest drawn on the radial (yellow), median (red) and ulnar (blue) nerves.

Figure S2. The columns show maps from unprocessed (blip-up and blip-down data) and pre-processed datasets. The rows contain maps of quantitative anisotropy (QA), fractional anisotropy (FA), mean diffusivity (MD) and the principal eigenvector (v1) with the colors red, green and blue representing diffusion in x, y and z directions.

Figure S3. Raincloud plot showing the Sørensen-Dice similarity coefficient for all datasets against the T1w reference images, stratified by the anatomical region.

Figure S4. Raincloud plot showing the change in Sørensen-Dice similarity coefficient scores after preprocessing (all pipelines are shown here) against the T1w reference. The horizontal line represents no (0%) change. The boxplots show the median improvement (0.014) was small (analogous to the geometric mean of ~1% stated in the main text) and the majority of the datapoints are clustered around minimal change (IQR -2%–4%). However, there are important outliers whereby preprocessing both improved and worsened the Sørensen-Dice similarity coefficient substantially (the worst 5% of processed slices had 0.16 lower Sørensen-Dice similarity coefficients, whilst the best 5% improved by 0.16. Overall, distortions in the arm and forearm were better corrected than data around the elbow, although the differences are not clinically meaningful.

Figure S5. A scatter plot with linear fit and Bland–Altman plot showing poor agreement for FA between datasets preprocessed with TOPUP & eddy versus DSI Studio. Pearson's $r = 0.360$, ICC = 0.075.

Figure S6. A scatter plot with linear fit and Bland–Altman plot showing poor agreement for FA between datasets preprocessed with TOPUP & eddy versus TORTOISE. Pearson's $r = 0.231$, ICC = 0.038.

Figure S7. A scatter plot with linear fit and Bland–Altman plot showing poor agreement for FA between datasets preprocessed with DSI Studio versus TORTOISE. Pearson's $r = 0.457$, ICC = 0.139.

Figure S8. A scatter plot with linear fit and Bland–Altman plot showing poor agreement for FA between datasets preprocessed with TOPUP versus DSI Studio. Pearson's $r = 0.536$, ICC = 0.172.

Figure S9. A scatter plot with linear fit and Bland–Altman plot showing poor agreement for FA between datasets

preprocessed with TOPUP versus TORTOISE. Pearson's $r = 0.2919$, ICC = 0.103.

Figure S10. A scatter plot with linear fit and Bland–Altman plot showing poor agreement for FA between datasets preprocessed with TOPUP versus TOPUP & eddy. Pearson's $r = 0.414$, ICC = 0.098.

Table S1. Mean DTI metrics of the median, ulnar and radial nerves for each pre-processing pipeline, categories by anatomical location.

How to cite this article: Wade RG, Tam W, Perumal A, et al. Comparison of distortion correction preprocessing pipelines for DTI in the upper limb. *Magn Reson Med.* 2023;1-11. doi: 10.1002/mrm.29881