

Sex-related Variability of White-Matter Tracts is Robust to Tractography Methodology

Matthew Amandola

`matthew.amandola.1@vanderbilt.edu`

Vanderbilt University Institute of Imaging Science

Bastien Herlin

Université Paris-Saclay, CNRS, CEA

Michael E. Kim

Vanderbilt University

Simon Vandekar

Vanderbilt University Medical Center

Ivy Uszynski

Université Paris-Saclay, CNRS, CEA

Bennett Landman

Vanderbilt University Institute of Imaging Science

Cyril Poupon

Université Paris-Saclay, CNRS, CEA

Kurt G. Schilling

Vanderbilt University Institute of Imaging Science

Short Report

Keywords: tractography, reproducibility, sex-effects, microstructure

Posted Date: April 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9439593/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Diffusion tractography is the prominent *in-vivo* technique to study and investigate white-matter pathways in the human brain. While tractography is a powerful method, recent work suggests that different tractography methods can produce strikingly different representations of the same white-matter pathway. This multitude of differing options and diverging pipelines makes tractography related group-effects difficult to generalize, as it is currently unclear whether group-level inferences made using one tractography pipeline can be expected to hold when a different, equally defensible pipeline is applied to the same data. Here, we test the generalizability of sex-related changes on tractography-derived features by analyzing the exact same datasets with two equally reasonable pipelines which differ in model fitting, tractography reconstruction, and microstructure and volumetric analysis.

We found that despite differences in analysis, the resulting patterns and biological interpretations of sex effects rarely disagreed across methods. Microstructural effects between methods were remarkably consistent between protocols, only displaying one significant disagreement out of 343 comparisons (.29%). However, discrepancies were more common among volumetric effects, displaying 24% significant disagreement. Moreover, we found that reconstruction methods are differentially sensitive to tractography-derived features, as bundles derived from targeted tractography were much more sensitive to volumetric effects than tractogram-based tractography, potentially explaining the volumetric discrepancy between methods.

This study indicates that reasonable methodological choices are unlikely to lead two investigators to fundamentally opposing conclusions about sex differences in white-matter, and that the robustness of tractography findings is similar to established fields of science. More broadly, this study presents an optimistic outlook on the future of tractography, as it provides an empirical benchmark for reproducibility and bolsters confidence in the generalizability and robustness of tractography-derived findings.

1. Introduction

Diffusion tractography allows investigators to study previously inaccessible white-matter pathways *in-vivo*, and has become the premier method for categorizing the neuroanatomical characteristics of human white-matter. Bundle segmentation using tractography has contributed to things like identifying new pathways (Catani et al., 2013), cross-validating histological findings (Amandola et al., 2026; Girard et al., 2020; Thiebaut de Schotten et al., 2011), and further understanding white-matter's role in cognition and disease (Forkel et al., 2022).

However, there is no consensus tractography processing pipeline, resulting in widely variable bundle segmentation protocols and techniques. As a result, recent studies have demonstrated that multiple different, yet equally reasonable, tractography methods can produce strikingly different representations of the same white-matter pathways (Maier-Hein et al., 2017; Schilling et al., 2021). This is due to the many different options at each step of the tractography workflow: reconstruction and model fitting (i.e., model choice and fitting strategy), tractography and bundle segmentation process (i.e., choice in fiber

orientation reconstruction, streamline propagation algorithm and bundle segmentation technique), and quantitative analysis (i.e., mean extraction and density-weighted extraction). Moreover, results from these varying techniques are compounded by the prevalence of false positives trajectories in tractography (Maier-Hein et al., 2017). As a result, it is unclear whether group-level inferences made using one tractography pipeline can be expected to hold when a different, equally defensible pipeline is applied to the same data.

While this is a crucial realization, this proposes another important question: do these variabilities between pipelines lead to contradictory interpretations associated with these pathways? To answer this, we were motivated by recent work utilizing tractography to study an exemplar biological effect: sex-related differences (Herlin et al., 2024). Our goal here is to test if the same study employing a different tractography pipeline would lead to identical conclusions, highlighting the importance of understanding the potential effects of tractography methodology on the generalizability of study results.

In the current study, we ask whether two different automated bundle segmentation methods applied to the same exact dataset lead to consistent conclusions about white-matter pathways. In direct collaboration with (Herlin et al., 2024), we directly tested the reproducibility and generalizability of tractography-derived sex differences in white-matter micro- and macrostructural features using two widely used automated tractography techniques. Both laboratories independently processed the Human Connectome Project Young Adult (HCP) (Van Essen et al., 2012) dataset using their own bundle segmentation methods, extracted a common set of microstructural features, and tested for sex-related differences. We then quantified agreement between pipelines in terms of statistical significance and effect directionality, asking whether two automated tractography methods lead to convergent versus contradictory conclusions. We found that despite substantial differences in the definitions, shapes, and spatial extents of the reconstructed bundles, both the resulting patterns of sex effects and, critically, the biological interpretation of microstructural effects were remarkably congruent, whereas volumetric effects are more likely to diverge. This work provides an optimistic and nuanced outlook on the robustness and interpretability of tractography-derived findings.

2. Methods

2.1 Dataset

Both labs used the exact same dataset: We analyzed 1,065 healthy young adults (575 women, 490 men; age range 22–35 years) from the Human Connectome Project Young Adult (HCP-YA) minimally preprocessed cohort (Glasser et al., 2013), which had performed intensity normalization of the mean b0 image, EPI distortion correction, EDDY correction, and gradient nonlinearity correction to the diffusion data. Differences in processing start to arise at: 1. dMRI processing, 2. Tractography reconstruction, and 3. Microstructural and volumetric analysis of the tracts (Fig. 1), detailed below.

2.1.1 Reconstruction and Model Fitting

Method 1 is an in-house pipeline designed by the Neurospin team (<https://framagit.org/coupon/gkg>) based on the Ginkgo toolbox. This toolbox calculates the Diffusion Tensor Imaging (DTI) model (Basser et al., 1994) and the Neurite Orientation Dispersion and Density Imaging (NODDI) model (H. Zhang et al., 2012) on the data. This toolbox also calculates the Orientation Distribution Functions (ODF) for each voxel using the Q-Ball model (Descoteaux et al., 2007). Both NODDI and ODF were calculated using all three diffusion shells. DTI metrics were calculated using only the $b = 1000 \text{ s/mm}^2$ shell, using weighted least squares (WLS).

Method 2 employs MRTrix3 (Tournier et al., 2019) to fit the DTI model on the data using all volumes with $b = 1000 \text{ s/mm}^2$, using iterated weighted least-squares (IWLS). The NODDI model is then fit to the data using the scilpy toolkit (Renauld et al., 2026) using all three diffusion shells. Method 2 uses constrained spherical deconvolution (CSD) (Jeurissen et al., 2014) to calculate ODFs, contained within the tractography process (see 2.2.2 - *Tractography Reconstruction*).

Both methods calculate fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) using their respective DTI-derived scalar map pipelines, as well as isotropic water fraction (ISOWF), neurite density index (NDI), and orientation dispersion index (ODI) using their respective NODDI pipelines.

2.1.2 Tractography Reconstruction

For tractography reconstruction, Method 1 uses ODF maps to conduct whole-brain probabilistic tractography, resulting in a complete tractogram for each subject (Perrin et al., 2005). Then, a predefined deep white-matter atlas (Chauvel et al., 2023; Herlin et al., 2023) is coregistered to each subject’s native diffusion space from standard MNI (Montreal Neurological Institute) (Fonov et al., 2011) using the Advanced Normalization Tools (ANTs) toolbox (Avants et al., 2008) to first create the native-to-MNI transformation, and then the inverse MNI-to-native transformation, which is then applied to the deep white-matter atlas. Automated bundle segmentation from the white-matter atlas is then performed in the native diffusion space using the subject’s tractogram. Streamlines are then assigned to a possible 77 different tracts based on the white-matter atlas using pairwise distance. For full details on the mechanics of this automatic segmentation algorithm, see Herlin et al., 2024.

Method 2 uses Tractseg (Wasserthal et al., 2018), which automatically segments anatomically defined deep white-matter pathways. Tractseg uses MRTrix3 (Tournier et al., 2019) to compute the ODF peaks per voxel using CSD. Using a convolutional neural network model, and then performs bundle-specific tractography based on these ODF peaks, and outputs 72 predefined bundles of streamlines.

We identified 49 white-matter bundles these methods have in common. For a full list of bundles included in this analysis, see Table 1.

Table 1 – List of tracts shared between pipelines

Abbreviation	Full Tract Name
CA	Anterior Commissure
CC_1	Corpus Callosum: Rostrum
CC_2	Corpus Callosum: Genu
CC_3	Corpus Callosum: Rostral Midbody
CC_4	Corpus Callosum: Anterior Midbody
CC_5	Corpus Callosum: Posterior Midbody
CC_6	Corpus Callosum: Isthmus
CC_7	Corpus Callosum: Splenium
MCP	Middle Cerebellar Peduncle
AF_left	Left Arcuate Fasciculus
AF_right	Right Arcuate Fasciculus
ATR_left	Left Anterior Thalamic Radiation
ATR_right	Right Anterior Thalamic Radiation
CG_left	Left Cingulum
CG_right	Right Cingulum
CST_left	Left Corticospinal Tract
CST_right	Right Corticospinal Tract
FX_left	Left Fornix
FX_right	Right Fornix
ICP_left	Left Inferior Cerebellar Peduncle
ICP_right	Right Inferior Cerebellar Peduncle
IFO_left	Left Inferior Fronto-Occipital Fasciculus
IFO_right	Right Inferior Fronto-Occipital Fasciculus
ILF_left	Left Inferior Longitudinal Fasciculus
ILF_right	Right Inferior Longitudinal Fasciculus
MLF_left	Left Middle Longitudinal Fasciculus
MLF_right	Right Middle Longitudinal Fasciculus
OR_left	Left Optic Radiation

OR_right	Right Optic Radiation
SCP_left	Left Superior Cerebellar Peduncle
SCP_right	Right Superior Cerebellar Peduncle
SLF_I_left	Left Superior Longitudinal Fasciculus I
SLF_I_right	Right Superior Longitudinal Fasciculus I
SLF_II_left	Left Superior Longitudinal Fasciculus II
SLF_II_right	Right Superior Longitudinal Fasciculus II
SLF_III_left	Left Superior Longitudinal Fasciculus III
SLF_III_right	Right Superior Longitudinal Fasciculus III
ST_OCC_left	Left Striato-Occipital Tract
ST_OCC_right	Right Striato-Occipital Tract
ST_PAR_left	Left Striato-Parietal Tract
ST_PAR_right	Right Striato-Parietal Tract
STR_left	Left Superior Thalamic Radiation
STR_right	Right Superior Thalamic Radiation
Striato_Central_Left	Left Striato-Central Tract
Striato_Central_Right	Right Striato-Central Tract
Striato_Frontal_Left	Left Striato-Frontal Tract
Striato_Frontal_Right	Right Striato-Frontal Tract
T_OCC_left	Left Thalamo-Occipital Tract
T_OCC_right	Right Thalamo-Occipital Tract
T_PAR_left	Left Thalamo-Parietal Tract
T_PAR_right	Right Thalamo-Parietal Tract
UF_left	Left Uncinate Fasciculus
UF_right	Right Uncinate Fasciculus

2.1.3 Microstructural and Volumetric Analysis

Method 1 computes total brain volume (TBV) using Freesurfer (Fischl, 2012). Volume for each tract is then computed with a streamline density-weighted mask with a threshold of 5 fibers per voxel. Normalized volume (N. Vol.) is then calculated by dividing each tract's volume by the corresponding

individual's TBV. Mean values for each tract and feature are then calculated after resampling the quantitative maps to 0.1mm via trilinear interpolation.

Method 2 also extracts streamline density-weighted averages of all path-feature pairs. For each tract, Method 1 uses the Scilpy toolkit (Renauld et al., 2026) to compute streamline density-weighted averages (note, in contrast to mean value within a binary mask) of DTI, NODDI, and volumetric features in native space. Similar to Method 1, tract volume was normalized by TBV, measured as estimated total intracranial volume and cerebral white-matter volume.

2.2 Statistical Analysis

2.2.1 Sex Differences in Path-Feature Pairs

Both methods extract the following features for each subject and tract: AD, FA, MD, RD, NDI, ODI, ISOWF, volume, and normalized volume. Sex differences were tested independently for every tract-feature combination using Student's t-tests comparing males and females, with Bonferroni correction applied to control for multiple comparisons across all tract-feature tests. With our 49 tracts and 9 features, there are a total of 4421 comparisons in this study, which necessitates a significance threshold of $p < .00001$ following Bonferroni correction. However, in the original study (Herlin et al., 2024), significance was tested using a threshold of $p < 0.000065$ as there were more tract-feature pairs due to the current paper focusing on the tracts the two methods have in common. In order to compare group correspondence to the original analysis, we employed this more conservative threshold. We then computed effect size magnitudes for sex differences within each tract-feature pair using Cohen's d test, with $|d| < 0.2$ meaning a negligible effect size, $|d| = 0.2 - 0.5$ meaning small effect size, $|d| = 0.5 - 0.8$ meaning medium effect size, and $|d| > 0.8$ meaning large effect size.

2.2.2 Agreement between Tractography Pipelines

We quantified agreement between the two methods on sex differences between tracts in features in their datasets. Agreement between methods was quantified at four levels, listed below:

- **Strict Agreement:** Both results statistically significant with the same direction, or neither results are significant
- **Soft Agreement:** Both results statistically significant with the same direction, or one significant with the other non-significant but same direction
- **Strict Disagreement:** Both results statistically significant with opposite directions
- **Soft Disagreement:** Both results statistically significant with opposite directions, or one significant with the other non-significant but opposite direction

Finally, to compare the sensitivity of the two methods, we employed a non parametric bootstrap with 10,000 iterations to estimate 95% confidence intervals (CI) for the difference in effect size magnitude

between techniques for each feature, and considered sensitivity to differ significantly when the CI excluded zero.

3. Results

3.1 Methodology correspondence

Across most features and tracts, both methods agreed that males show higher values; FA was the sole exception, with both methods consistently indicating higher FA in females across tracts. To evaluate the generalizability of the sex effects of tractography-derived features, we calculated agreement and disagreement between the two methods. Across all 441 tract-feature comparisons, we observed strong correspondence between the two automated tractography techniques (Fig. 2). Between the two methodologies, 61.8% of tract-feature pairs reached strict agreement, as both pipelines yield either a significant sex effect in the same direction or a non-significant result. When the criterion was expanded to soft agreement, agreement increased to 86.7%, as the direction of the effect, irrespective of whether each pipeline's result passed multiple comparison correction, was consistent between the two methodologies. Thus, in the vast majority of cases, both techniques pointed toward the same qualitative pattern of sex differences, even when only one pipeline reached formal significance.

Direct contradictions between techniques were rare. Strict disagreement occurred in only 2% of comparisons: out of 441 tract-feature pairs, there were 9 instances in which both methods produced statistically significant but directionally opposing effects. When expanded to soft disagreement to include any opposite directionality, regardless of significance, disagreement was 24.2%. These disagreements were typically characterized by very small effect sizes near zero in at least one pipeline, consistent with statistical noise rather than systematic reversals of effect. These results suggest that correspondence between tractography methodologies, regardless of reconstruction technique, algorithm, and feature extraction was exceptionally high.

Notably, the disagreement between methods was driven by volumetric features (Fig. 3). Specifically, while tract-volume sex-effects were fully in agreement, normalized volume effects frequently diverged. Microstructural features very rarely disagreed: out of 343 comparisons, only one tract-feature was significantly different between methods (.3%). However, when considering only volumetric effects, strict disagreement rose to 8.1%, with 8 tract-feature pairs significantly different between the two methods.

3.2 Methodological Sensitivity

We also examined whether one method was systematically more sensitive to sex differences in particular features (Fig. 4). In this case, sensitivity corresponds to a significantly different effect size between methods revealed by bootstrapping. For microstructural measures, the bootstrap analysis of effect size magnitude revealed that with the exception of NDI, there was no differentiable pattern in feature sensitivity between the two methods, with Method 1 significantly more sensitive in 42 features,

and Method 2 more sensitive in 48 features. When considering NDI, Method 2 was significantly more sensitive, making the totals 43 for Method 1 and 83 for Method 2, as NDI alone contributed 35 additional tract-feature pairs for Method 2

For volumetric effects, Method 2 was considerably more sensitive, as it was significantly more sensitive in 65 tract-feature pairs, compared to 3 tract-feature pairs for Method 1. This difference in sensitivity may partly explain the considerable divergence in sex effects between the two methods, as normalized volume displayed the most divergence between all the features.

Taken together, these results indicate that while tractography pipelines can differ in how often they detect statistically significant sex effects for particular features, the qualitative conclusions drawn about the presence and direction of sex differences in white-matter microstructure are consistent across methods when applied to the same high quality dataset, though volumetric effects are still seem quite sensitive to methodology.

4. Discussion

Correspondence between tractography methods is critical for interpreting diffusion MRI studies that rely on automated bundle reconstructions. It is crucial that investigators understand whether observed group differences, such as sex effects, are robust properties of white-matter organization or artifacts of a particular processing pipeline. This issue is especially salient for tractography, which remains vulnerable to spurious streamlines and lacks an absolute anatomical ground truth (Dyrby et al., 2007). Indeed, tractography is a complex, multifaceted analysis, where segmented bundles are impacted by each step of the processing pipeline (Maier-Hein et al., 2017), highlighting the significance in understanding pipeline variability's effect on potential biological interpretations.

In this large sample of HCP-YA participants, we found encouraging evidence that tractography-derived sex differences in white-matter microstructure are generalizable across two independent, state of the art automated pipelines. Importantly, biological interpretation remained consistent between methodologies, with strict agreement in both significance and directionality observed for roughly two thirds of tract-feature comparisons, a rate comparable to (Brodeur et al., 2026), and in some cases exceeding (Tyner et al., 2026), cross-pipeline reproducibility reported in other scientific fields. Additionally, overall agreement in the direction of effect was 90%, with only one strict microstructural disagreement out of 343 possible comparisons. Encouragingly, while the motivator of this study was work by Herlin and colleagues (2024), it is important to note that these results are in consensus and are further supported by previous sex-related tractography literature, which typically finds that women exhibit higher FA in these long-range bundles (Herlin et al., 2024; Kochunov et al., 2015; Raikes et al., 2025). Notably, this includes recent work using altogether different methodology by employing multitensor tractography and using unweighted mean metric values (F. Zhang et al., 2025), and found congruent sex-related findings in tracts such as the arcuate fasciculus, cerebellar peduncles, corticospinal tract, and the thalamic radiations.

While microstructural effects were remarkably congruent, our findings suggest that volumetric features were sensitive to methodology, comprising 8 of the 9 strict disagreements between methods, and suggesting that women may show a higher volume of white-matter compared to gray-matter, as this was true across bundles. This may reflect the inconsistent nature of sex-related effects of white-matter volume in past literature, as in the past, studies report that men display higher white-matter to gray-matter ratio (Kanaan et al., 2012). However, more recent papers correcting for TBV suggest that there is no differentiable trend in white-matter volume between men and women (Ramzanpour et al., 2023; Muer et al., 2024), suggesting our findings may stem from the given reconstruction technique. These differences may also be explained by methodological differences in sensitivity to volumetric features: Method 2 was significantly more sensitive in volume, normalized volume, and NDI, where virtually all disagreements between methods arose. Therefore, researchers should consider which technique they use when probing specific pathways or features, particularly if their primary goal is maximizing sensitivity for a given metric.

While the current study focuses on sex-effects, we believe our results are overall optimistic for tractography, and are comparable to many scientific fields as a whole: they indicate that reasonable methodological choices may not always lead to perfect replication of results, but that they are unlikely to lead two investigators to fundamentally opposing conclusions about sex differences in white-matter microstructure. More broadly, this work provides an empirical benchmark for cross-pipeline reproducibility and gives tractography researchers greater confidence in the generalizability and robustness of their findings.

Declarations

Funding declaration: Dr. Matthew Amandola is funded by National Institute of Health (NIH) T32 EB001628; Dr. Bastien Herlin, Dr. Ivy Uszynski, and Dr. Cyril Poupon are funded by the European Union's Horizon 2020 Framework Program for Research and Innovation, Grant No. 945539 (Human Brain Project SGA3); Dr. Simon Vandekar is funded by NIH R01 MH123563; Dr. Bennett Landman is funded by NIH R01 EB017230; Dr. Kurt Schilling is funded by NIH K01 EB032898.

References

1. Amandola, M., Kim, M. E., Rheault, F., Landman, B., & Schilling, K. (2026). Bridging Histology and Tractography: First In Vivo Visualization of Short-Range Prefrontal Connections Informed by Primate Tract-Tracing. *Human Brain Mapping*, 47(5), e70520. <https://doi.org/10.1002/hbm.70520>
2. Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. <https://doi.org/10.1016/j.media.2007.06.004>
3. Basser, P. J., Mattiello, J., & LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1), 259–267. [https://doi.org/10.1016/S0006-3495\(94\)80775-1](https://doi.org/10.1016/S0006-3495(94)80775-1)

4. Bookheimer, S. Y., Salat, D. H., Terpstra, M., Ances, B. M., Barch, D. M., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Diaz-Santos, M., Elam, J. S., Fischl, B., Greve, D. N., Hagy, H. A., Harms, M. P., Hatch, O. M., Hedden, T., Hodge, C., Japardi, K. C., Kuhn, T. P., ... Yacoub, E. (2019). The Lifespan Human Connectome Project in Aging: An overview. *NeuroImage*, 185, 335–348. <https://doi.org/10.1016/j.neuroimage.2018.10.009>
5. Brodeur, A., Mikola, D., Cook, N., Fiala, L., Brailey, T., Briggs, R., De Gendre, A., Dupraz, Y., Gabani, J., Gauriot, R., Haddad, J., Lima, G., Ankel-Peters, J., Dreber, A., Campbell, D., Kattan, L., Marino Fages, D., Mierisch, F., Sun, P., ... Zhong, Y. (2026). Reproducibility and robustness of economics and political science research. *Nature*, 652(8108), 151–156. <https://doi.org/10.1038/s41586-026-10251-x>
6. Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., Thompson, C. K., Thiebaut de Schotten, M., Dell'Acqua, F., Weintraub, S., & Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain: A Journal of Neurology*, 136(Pt 8), 2619–2628. <https://doi.org/10.1093/brain/awt163>
7. Chauvel, M., Uszynski, I., Herlin, B., Popov, A., Leprince, Y., Mangin, J.-F., Hopkins, W. D., & Poupon, C. (2023). In vivo mapping of the deep and superficial white matter connectivity in the chimpanzee brain. *NeuroImage*, 282, 120362. <https://doi.org/10.1016/j.neuroimage.2023.120362>
8. Descoteaux, M., Angelino, E., Fitzgibbons, S., & Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3), 497–510. <https://doi.org/10.1002/mrm.21277>
9. Dyrby, T. B., Sogaard, L. V., Parker, G. J., Alexander, D. C., Lind, N. M., Baaré, W. F. C., Hay-Schmidt, A., Eriksen, N., Pakkenberg, B., Paulson, O. B., & Jelsing, J. (2007). Validation of in vitro probabilistic tractography. *NeuroImage*, 37(4), 1267–1277. <https://doi.org/10.1016/j.neuroimage.2007.06.022>
10. Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
11. Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., & Brain Development Cooperative Group. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1), 313–327. <https://doi.org/10.1016/j.neuroimage.2010.07.033>
12. Forkel, S. J., Friedrich, P., Thiebaut de Schotten, M., & Howells, H. (2022). White matter variability, cognition, and disorders: A systematic review. *Brain Structure & Function*, 227(2), 529–544. <https://doi.org/10.1007/s00429-021-02382-w>
13. Girard, G., Caminiti, R., Battaglia-Mayer, A., St-Onge, E., Ambrosen, K. S., Eskildsen, S. F., Krug, K., Dyrby, T. B., Descoteaux, M., Thiran, J.-P., & Innocenti, G. M. (2020). On the cortical connectivity in the macaque brain: A comparison of diffusion tractography and histological tracing data. *NeuroImage*, 221, 117201. <https://doi.org/10.1016/j.neuroimage.2020.117201>
14. Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., Jenkinson, M., & WU-Minn HCP Consortium.

- (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>
15. Herlin, B., Uszynski, I., Chauvel, M., Dupont, S., & Poupon, C. (2024). Sex-related variability of white matter tracts in the whole HCP cohort. *Brain Structure & Function*, 229(7), 1713–1735. <https://doi.org/10.1007/s00429-024-02833-0>
 16. Herlin, B., Uszynski, I., Chauvel, M., Poupon, C., & Dupont, S. (2023). Cross-subject variability of the optic radiation anatomy in a cohort of 1065 healthy subjects. *Surgical and Radiologic Anatomy: SRA*, 45(7), 849–858. <https://doi.org/10.1007/s00276-023-03161-4>
 17. Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., & Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, 103, 411–426. <https://doi.org/10.1016/j.neuroimage.2014.07.061>
 18. Kanaan, R. A., Allin, M., Picchioni, M., Barker, G. J., Daly, E., Shergill, S. S., Woolley, J., & McGuire, P. K. (2012). Gender Differences in White Matter Microstructure. *PLoS ONE*, 7(6), e38272. <https://doi.org/10.1371/journal.pone.0038272>
 19. Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T. E., Wright, S. N., Hong, L. E., Patel, B., Behrens, T., Jbabdi, S., Andersson, J., Lenglet, C., Yacoub, E., Moeller, S., Auerbach, E., Ugurbil, K., Sotiropoulos, S. N., Brouwer, R. M., ... Van Essen, D. C. (2015). Heritability of fractional anisotropy in human white matter: A comparison of Human Connectome Project and ENIGMA-DTI data. *NeuroImage*, 111, 300–311. <https://doi.org/10.1016/j.neuroimage.2015.02.050>
 20. Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., He, R., Li, Q., ... Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8(1), 1349. <https://doi.org/10.1038/s41467-017-01285-x>
 21. Muer, J. D., Didier, K. D., Wannebo, B. M., Sanchez, S., Khademi Motlagh, H., Haley, T. L., Carter, K. J., Banks, N. F., Eldridge, M. W., Serlin, R. C., Wieben, O., & Schrage, W. G. (2024). Sex differences in grey matter, white matter, and regional brain perfusion in young, healthy adults. *American Journal of Physiology-Heart and Circulatory Physiology*, ajpheart.00341.2024. <https://doi.org/10.1152/ajpheart.00341.2024>
 22. Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Rivière, D., Constantinesco, A., Le Bihan, D., & Mangin, J.-F. (2005). Fiber Tracking in q-Ball Fields Using Regularized Particle Trajectories. In G. E. Christensen & M. Sonka (Eds.), *Information Processing in Medical Imaging* (Vol. 3565, pp. 52–63). Springer Berlin Heidelberg. https://doi.org/10.1007/11505730_5
 23. Raikes, A. C., Dyke, J. P., Nerattini, M., Boneu, C., Ajila, T., Fauci, F., Battista, M., Pahlajani, S., Williams, S., Brinton, R. D., & Mosconi, L. (2025). White matter microstructural and macrostructural profiles during midlife reveal sex differences between men and women at different menopausal stages. *Scientific Reports*, 15(1), 40312. <https://doi.org/10.1038/s41598-025-24136-y>

24. Ramzanpour, M., Jafari, B., Smith, J., Allen, J., & Hajiaghamemar, M. (2023). Comprehensive study of sex-based anatomical variations of human brain and development of sex-specific brain templates. *Brain Multiphysics*, 4, 100077. <https://doi.org/10.1016/j.brain.2023.100077>
25. Renaud, E., Boré, A., Poirier, C., Valcourt-Caron, A., Karan, P., Théberge, A., Théaud, G., Edde, M., Poulin, P., Girard, G., Houde, J.-C., Gagnon, A., St-Onge, E., Little, G.,
26. Legarreta, J. H., Thoumyre, S., Grenier, G., El Yamani, Z., Ocampo Pineda, M., ... Descoteaux, M. (2026). Tractography analysis with the scilpy toolbox. *Aperture Neuro*, 6, 5.
27. Schilling, K. G., Rheault, F., Petit, L., Hansen, C. B., Nath, V., Yeh, F.-C., Girard, G., Barakovic, M., Rafael-Patino, J., Yu, T., Fisch-Gomez, E., Pizzolato, M., Ocampo-Pineda, M., Schiavi, S., Canales-Rodríguez, E. J., Daducci, A., Granziera, C., Innocenti, G., Thiran, J.-P., ... Descoteaux, M. (2021). Tractography dissection variability: What happens when 42 groups dissect 14 white matter bundles on the same dataset? *NeuroImage*, 243, 118502. <https://doi.org/10.1016/j.neuroimage.2021.118502>
28. Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., & Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, 14(10), 1245–1246. <https://doi.org/10.1038/nn.2905>
29. Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202, 116137. <https://doi.org/10.1016/j.neuroimage.2019.116137>
30. Tyner, A. H., Abatayo, A. L., Daley, M., Field, S., Fox, N., Haber, N. A., Hahn, K. M., Struhl, M. K., Mawhinney, B., Miske, O., Silverstein, P., Soderberg, C. K., Stankov, T., Abbasi, A., Aberson, C. L., Aczel, B., Adamkovič, M., Albayrak, N., Allen, P. J., ... Errington, T. M. (2026). Investigating the replicability of the social and behavioural sciences. *Nature*, 652(8108), 143–150. <https://doi.org/10.1038/s41586-025-10078-y>
31. Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., Della Penna, S., Feinberg, D., Glasser, M. F., Harel, N., Heath, A. C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., ... Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, 62(4), 2222–2231. <https://doi.org/10.1016/j.neuroimage.2012.02.018>
32. Wasserthal, J., Neher, P., & Maier-Hein, K. H. (2018). TractSeg—Fast and accurate white matter tract segmentation. *NeuroImage*, 183, 239–253. <https://doi.org/10.1016/j.neuroimage.2018.07.070>
33. Zhang, F., Rushmore, J., Li, Y., Cetin-Karayumak, S., Song, Y., Cai, W., Westin, C.-F., Levitt, J. J., Makris, N., Rathi, Y., & O'Donnell, L. J. (2025). Study of Sex Differences in the Whole Brain White Matter Using Diffusion MRI Tractography and Suprathreshold Fiber Cluster Statistics. *bioRxiv: The Preprint Server for Biology*, 2025.09.27.679006. <https://doi.org/10.1101/2025.09.27.679006>
34. 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>

Figures

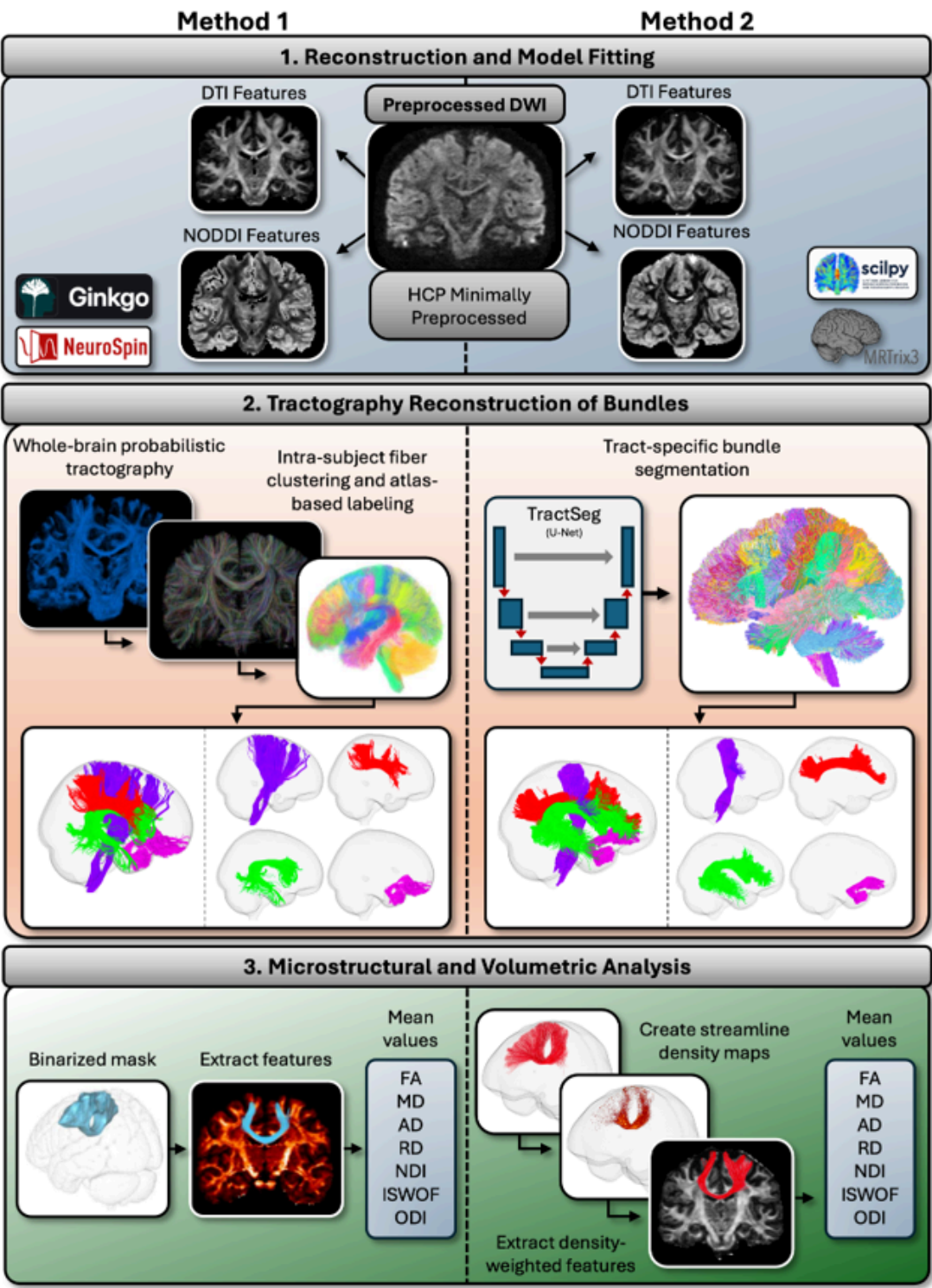


Figure 1

Pipeline overview between the whole-brain tractogram method (Method 1) vs the bundle-specific bundle segmentation method (Method 2).

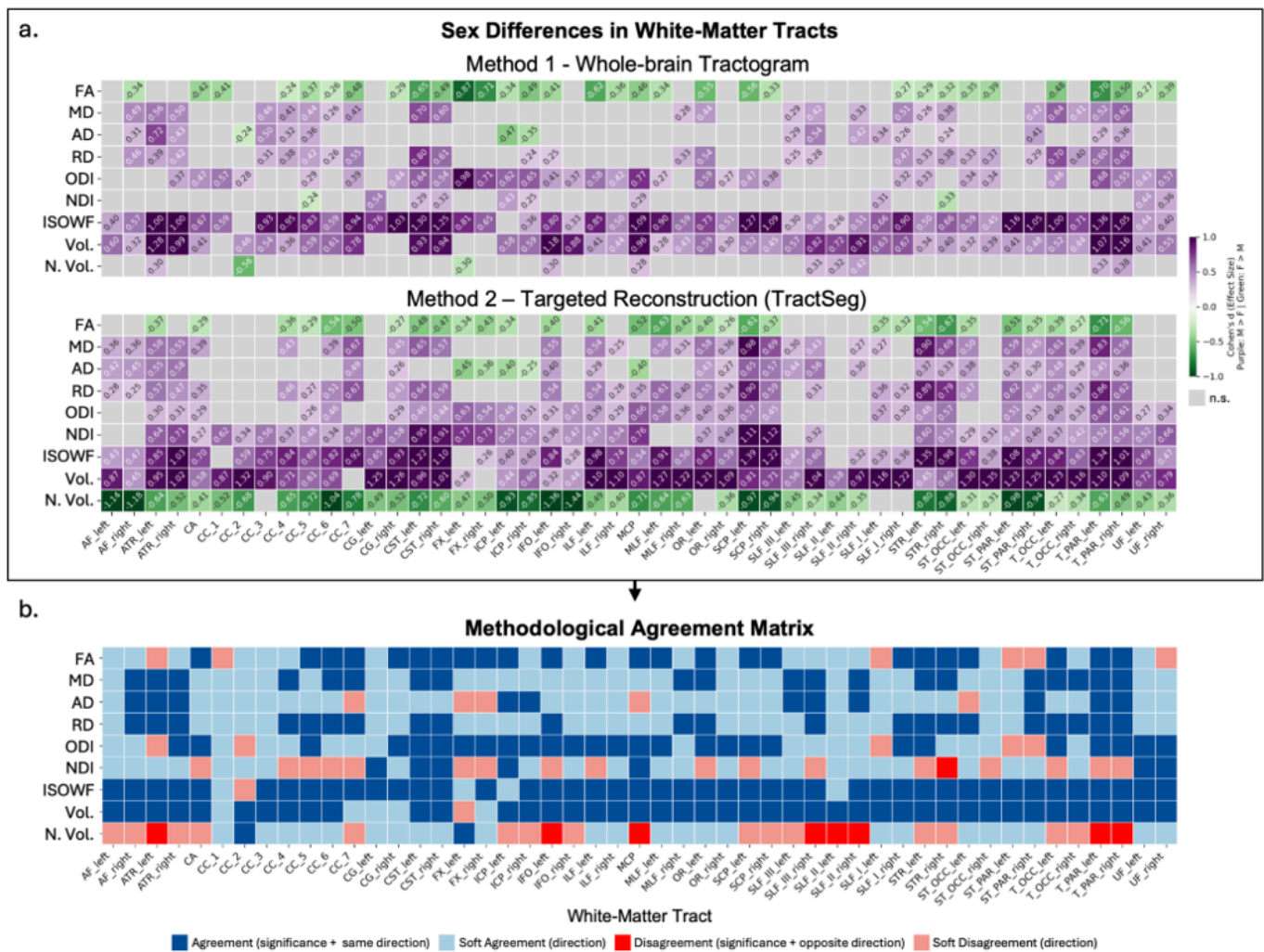


Figure 2

Methodological correspondence. a.) Sex differences in tractography-derived metrics calculated from Method 1 (top) and Method 2 (bottom). Purple corresponds to significantly higher in males, green corresponds to significantly higher in females, and hue corresponds to Cohen's d. Grey corresponds to nonsignificant. b.) Overall correspondence between methods. Dark blue = strict agreement; light blue = soft agreement; dark red = strict disagreement; light red = soft disagreement. Vol. = Volume; N. Vol. = Volume normalized by total brain volume.

Microstructural vs. Volumetric Feature-Based Agreement

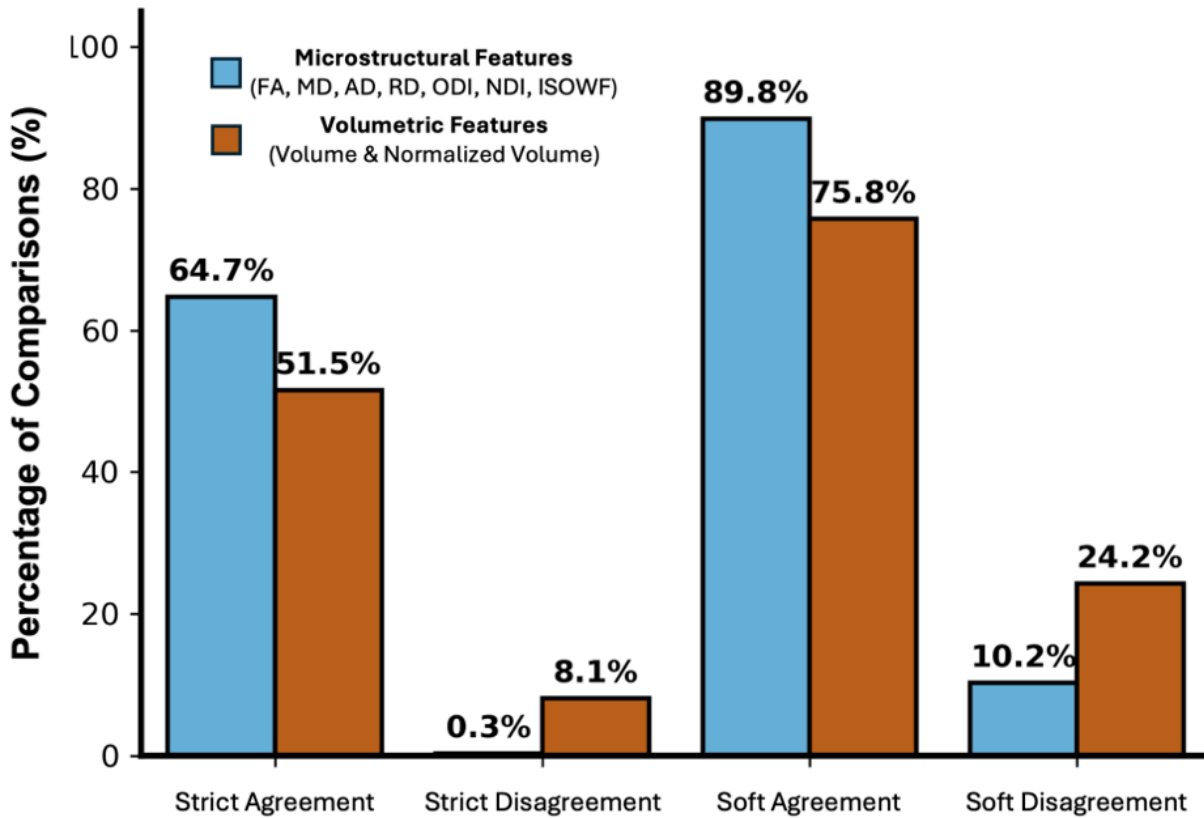


Figure 3

Percentage of methodological correspondence in microstructural features (blue) vs. volumetric features (red). Overall disagreement was driven mostly by volumetric features, comprising 8 out of 9 total strict disagreements.

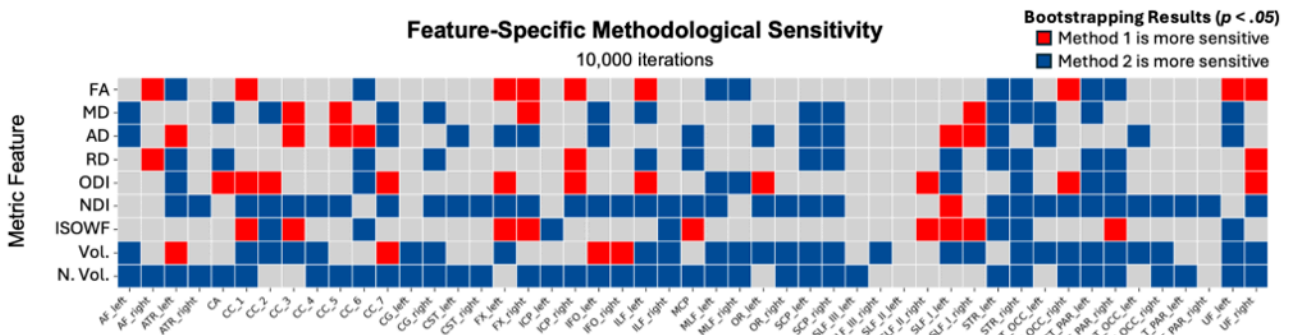


Figure 4

Feature-Specific Methodological Sensitivity. Bootstrap analysis for 10,000 iterations was conducted to calculate 95% confidence intervals in order to test if one method was significantly more sensitive for a

given metric feature than another. Orange = method 1 (whole-brain tractogram) is more sensitive; blue = method 2 (Tractseg) is more sensitive.