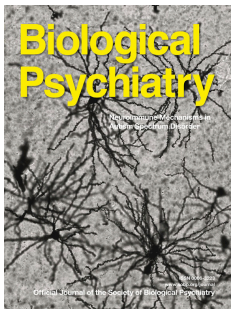
Journal Pre-proof



Lifespan Normative Models of White Matter Fractional Anisotropy: Applications to Early Psychosis

Ramona Cirstian, Natalie J. Forde, Gary Zhang, Gerhard S. Hellemann, Christian F. Beckmann, Nina V. Kraguljac, Andre F. Marquand

PII: S0006-3223(25)01379-4

DOI: https://doi.org/10.1016/j.biopsych.2025.07.021

Reference: BPS 15892

To appear in: *Biological Psychiatry*

Received Date: 10 February 2025

Revised Date: 11 July 2025

Accepted Date: 18 July 2025

Please cite this article as: Cirstian R., Forde N.J., Zhang G., Hellemann G.S., Beckmann C.F., Kraguljac N.V. & Marquand A.F., Lifespan Normative Models of White Matter Fractional Anisotropy: Applications to Early Psychosis, *Biological Psychiatry* (2025), doi: https://doi.org/10.1016/j.biopsych.2025.07.021.

Title

Lifespan Normative Models of White Matter Fractional Anisotropy: Applications to Early Psychosis

Short title

Lifespan Normative Models of White Matter Integrity

Authors

Ramona Cirstian*[1,2], Natalie J. Forde[1,2], Gary Zhang[3], Gerhard S. Hellemann[4], Christian F. Beckmann[1,2,5], Nina V. Kraguljac[6], Andre F. Marquand[1,2,7]

* ramona.cirstian@donders.ru.nl

Affiliations

1. Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

2. Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, the Netherlands

3. Department of Computer Science, University College London, London, UK

4. Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

5. Wellcome Centre for Integrative Neuroimaging - Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, UK

6. Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham, AL, USA

7. Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, UK

Abstract

Background

This study presents large-scale normative models of white matter (WM) organization across the lifespan, using diffusion MRI data from over 25,000 healthy individuals aged 0-100 years. These models capture lifespan trajectories and inter-individual variation in fractional anisotropy (FA), a marker of white matter integrity.

Methods

By addressing non-Gaussian data distributions, self-reported race, and site effects, the models offer reference baselines across diverse ages, and scanning conditions. We applied these FA models to the HCP Early Psychosis cohort and performed a multivariate analysis to map symptoms onto deviations from multimodal normative models using multi-view sparse canonical correlation analysis (msCCA).

Results

Our results reveal extensive white matter heterogeneity in psychosis, which is not captured by group-level analyses, with key regions identified, including the right uncinate fasciculus and thalami.

Conclusion

These normative models offer valuable tools for individualized WM deviation identification, improving precision in psychiatric assessments. All models are publicly available for community use.

Keywords

Normative models, White matter integrity, Psychosis, Lifespan trajectories, Diffusion MRI, Precision medicine

MAIN TEXT

Introduction

Over the past century, normative growth charts have become integral to paediatric practice, providing essential benchmarks for comparing individual growth patterns (height, weight, head circumference) with established population standards. These charts have facilitated a better understanding of typical developmental trajectories and have been crucial in identifying deviations from expected growth patterns which are used in clinical practice to determine if additional medical workup or treatment is required (1). This concept has recently been extended to the field of neuroimaging, where it allows for detailed, individual-level insights into lifespan trajectories of brain measures. By comparing individual neuroimaging data against large, normative reference datasets, researchers and clinicians can gain a deeper understanding of both typical and atypical brain development and aging (2–5).

In psychiatric disorders, traditional case-control studies have been valuable for detecting abnormalities in structural, microstructural, functional and neurometabolic brain signatures in patient groups compared to control groups. However, group comparisons are not designed to capture inter-individual heterogeneity which is prominent at the phenotypic and biological levels in virtually all psychiatric disorders. This significant translational gap hampers identification of specific biological markers that explain clinical heterogeneity in these disorders such as disease risk, severity, and progression, as well as responsiveness to pharmacological and non-pharmacological treatments and overall clinical outcomes. Normative modelling provides a precision framework that has emerged as a promising tool in this endeavour (6–8). By comparing brain imaging data against large reference cohorts, this method allows us to quantify deviations from expected norms at the individual level. It is now possible to capture deviation profiles in a single patient, which offers a more nuanced understanding of biological variations in psychiatric disorders. Even more importantly, it also has promises for bridging this translational gap by providing a foundational framework for developing tailored tools that capture disease risk and progression, as well as precision treatments tailored to individual brain pathology. For instance, normative models capture inter-individual biological variations that provided important insights into heterogeneity in schizophrenia, major depressive disorder, bipolar disorder, ADHD and autism spectrum disorders (6,9,10). Moreover, we have demonstrated that normative measures frequently outperform raw measures (e.g. cortical thickness in mm) in group difference testing, disease classification (11) and treatment response prediction (12).

We and others have created large-scale normative models that leveraged >50,000 healthy volunteer imaging datasets for structural (4,5,13) and functional MRI (11,14). To date, no large-scale normative models for diffusion MRI exist, due to its later adoption, high processing demands, and strong sensitivity to scanner and acquisition differences. Fractional Anisotropy (FA) is particularly well-suited for normative modelling of white matter, as it reflects directionally constrained diffusion influenced by axonal integrity, fibre coherence, and myelination—features that evolve across the lifespan and are disrupted in psychiatric disorders such as psychosis (15,16). Compared to other diffusion metrics such as mean diffusivity (MD), FA more directly captures tract-specific microstructural organization and is less sensitive to partial volume effects, making it a robust and interpretable marker for detecting spatially specific white matter deviations (17,18). To date, only two preliminary studies have fit normative models to diffusion weighted data (19),(20). In (19), the authors used approximately 1,300 single-shell DTI datasets collected at eight different sites using the same vendor to test performance of different statistical methods

within the normative framework and in (20) the authors focus principally on generating reference curves for data harmonisation.

The aims of this study are to: (i) develop normative models of FA, the most widely used diffusion metric in neuroimaging (21), across major white matter tracts using a large dataset of over 25,000 healthy individuals across a broad age range. By using high-quality diffusion MRI data from the UK Biobank and the Human Connectome Project, we seek to establish robust models that capture lifespan trajectories of white matter organization; (ii) investigate white matter FA in early psychosis, a prototypical psychiatric disorder that is known to be highly heterogeneous in disease severity and course, as well as clinical symptom expression and clinical outcomes. Using the HCP Early Psychosis (HCP-EP) dataset (22), we aim to map both group level differences and individual deviations from the normative model in order to better understand individual variability in white matter integrity; (iii) we aim to illustrate the value of normative models for multi-modal data fusion, by combining FA deviations with cortical thickness and subcortical brain volume deviations with the goal to identify multi-modal biological signatures and specific white matter pathways in psychosis associated with different psychosis symptom domains. Finally, (iv) we release all models freely to the community via our existing open-source software platforms (23).

Methods and materials

Data acquisition and processing

The construction of the lifespan dataset involved integrating data from five cohorts having high-quality multi-shell diffusion data, i.e.: the HCP Baby (24), HCP Development (25), HCP Young Adult (26), HCP Aging (27) datasets, and the UK Biobank (28). The demographic information is available in supplementary table 1 and 2 while an overview of the diffusion MRI acquisition schemes across datasets is available in supplementary table 3.

The processing of these datasets followed harmonized FSL-based pipelines, summarized in Figure 1A. Initially, pre-processing was performed: B0 intensity normalization, correction for EPI distortions, eddy-current-induced and movement corrections. These corrections were executed using the HCP-pipeline (29) for the HCP datasets while the UKB dataset was already processed according to the UKB documentation (30). Subsequently, we estimated the DTI model using DTIfit on the lowest shell value in order to extract the fractional anisotropy (FA) values. Following this, we ran Tract-Based Spatial Statistics (TBSS) (31) on the FA images which included registration to a standard space (FMRIB58_FA), projection of each individual's FA image to the standard space skeletonized image (threshold at 0.2) to generate skeletonized FA images for each individual in the same space. Finally, segmentation was conducted using the Johns Hopkins University (JHU) atlas (32). This process delineated 48 white matter (WM) tracts (listed the supplementary figure 4), for which we computed the mean FA values along the skeleton of each tract. To ensure consistency with UK Biobank data, we adopted the HCP preprocessing pipeline, enabling scalable and reproducible preprocessing across datasets. Diffusion MRI quality was evaluated using tract-wise normative modelling of FA, with extreme outliers ($|Z| > 4$) reviewed via visual inspection. Full details are provided in the Supplementary text: Quality control.

Normative modeling

To prepare for the modelling stage, we began by splitting the dataset of subjects (N=24,915) into two equal groups: a test set (N=12,457) and a training set (N=12,457), stratified to ensure an even distribution of sex, self-reported race, dataset and site. A normative model was then fit to the training set for each white matter tract. The model incorporated several covariates, including sex, age, and dummy coded race, and site. To address potential non-linear effects and non-Gaussian distributions, we employed a warped Bayesian linear regression (BLR) model and used in previous research (4,33). This approach involved applying a third-order polynomial B-spline basis expansion over age, with five evenly spaced knots, combined with a SinhArcsinh warping function. We chose to model site effects using a fixed effects framework, in line with prior work (4,11,33). We acknowledge that harmonisation techniques have also been employed for this purpose (34) and could have been applied here. However, we prefer to model the site effects explicitly within the normative model, which enables quantification and visualisation of the site effects and can be beneficial because harmonisation techniques can remove biological variation or induce bias if not employed correctly (35–37).

Next, we estimated deviation scores for each subject and white matter tract. In line with our prior work (38) we refit the models after excluding gross outliers having deviations larger than 4 standard deviations from the mean (Figure 1C). Once the models were refit with the cleaned data, we calculated the fit statistics, including explained variance, skew, and kurtosis. The extent of deviation for each subject was visualized by plotting individual z-scores against the mean and centiles of variation predicted by the model. All statistical analyses were conducted using Python version 3.8, with the Predictive Clinical Neuroscience PCN toolkit (GitHub, PCNtoolkit).

Application to a clinical dataset

Next, we applied the model to the Human Connectome Project Early Psychosis (HCP-EP) dataset (22), which includes multi-shell diffusion data and T1-weighted structural MRI derived from participants diagnosed with early psychosis (n=118) and control participants (n=55). The dataset's demographic distribution comprises 37% females and 63% males, with a racial composition of 58% White, 28% Black, 9% Asian, 1% Mixed, and 3% Other. Participants with early psychosis were diagnosed using the Structured Clinical Interview for DSM-5 (SCID-5) (39) and symptoms assessed with the Positive and Negative Syndrome Scale (PANSS) (40), including negative symptoms (e.g., social withdrawal), positive symptoms (e.g., hallucinations), disorganisation, and general psychopathology. The item-level data were subsequently summarized by the HCP-EP consortium using a standard factor model (41) and the positive, negative and cognitive symptom domain scores were used in in addition to the PANSS total score to quantify symptomatology across multiple domains (41). Medication status was also documented, including antipsychotic type and dosage converted to chlorpromazine equivalents.

The diffusion data were processed with the same pipeline as described above (Figure 1A), and structural data were processed using Freesurfer version 6.0 following similar procedures as we have described previously (4). Next, we divided this dataset into a training set, consisting of half of the control participants, and combined it with the larger training set described above to retrain the normative models for each white matter tract. Using transfer learning, as in our previous work, we can efficiently adapt the models with only a small amount of calibration data to account for site-specific effect. We then computed z-scores for the patients and remaining controls for the FA data and computed the deviations for cortical thickness and subcortical volumes derived from models we have previously

brought online (4). Note that the splits for this analysis were matched so that the same participants were in the training and test sets for diffusion and structural measures at each iteration.

We next assessed the mean difference of the deviations between patients and controls for each tract using a t-test with false discovery rate (FDR) correction for multiple testing (42). We then tested whether the proportion of extreme deviations differ between groups for each tract. To achieve this, we calculated the percentage of participants falling below and above the threshold in each of the 48 tracts. To achieve this, we set a z-score threshold between -2.6 and 2.6, which correspond to a p-=value of 0.01 as in prior work to identify extreme deviations then employed a non-parametric Mann-Whitney U test (43), again followed by FDR correction for multiple comparisons. This stringent threshold enhances the detection of significant deviations while controlling for false positives

We conducted a multimodal analysis to link psychiatric symptoms with brain deviations by combining diffusion and structural imaging data using multi-view sparse canonical correlation analysis (msCCA) [48]. This method identifies shared patterns between symptom scores and brain features while keeping structural and diffusion data separate, enhancing interpretability. Rather than tuning model parameters, we used fixed sparsity settings to prioritise reproducibility and reduce variability across runs. To ensure robustness, we applied stability selection across many random data splits, retaining only features that consistently contributed to the model. Generalisability was further tested through extensive permutation testing, where the symptom data were randomly shuffled to create a null distribution of associations. This allowed us to determine whether the observed brain–symptom relationships were stronger than expected by chance. For full methodological details see Supplementary text: Multimodal Analysis.

Results

Normative modelling

First, we assembled high-quality multi-shell diffusion data from five cohorts having closely matched acquisition and processing pipelines. A summary of the sample and processing is provided in Figure 1 with further details in the methods. We then fit lifespan normative models to these data on the basis of age, sex, site and self-reported race using warped Bayesian linear regression (BLR) and a non-linear basis expansion over age, in line with our prior work (4,33). We assessed the quality of the normative modeling fit using three key out-of-sample metrics, namely explained variance (EV), evaluating the fit of the median regression line, in addition to skewness and kurtosis, which evaluate the shape of the distribution used to model the centiles. These metrics offer insight into how well the models capture the underlying distribution of the data across 48 white matter tracts. The mean (standard deviation) EV was 0.37 (0.10), indicating good fit across different models. Skewness, and kurtosis were respectively -0.09 (0.12) and 0.42 (0.27), which together indicate that the shape was also appropriate for the data. Supplementary figure 1 shows a histogram of the EV, skew and kurtosis of the models.

We illustrate the trajectory and fit centiles for a selection of white matter tracts across the lifespan in Figure 2. The complete set can be found in the supplementary figure 2. In addition, we also show the results of models that do not include race in the supplementary figure 3.

Application to a clinical dataset

Next, we used these models to understand heterogeneity in white matter FA in psychosis. To achieve this, we applied these reference models to the HCP early psychosis (HCP-EP) dataset (N=173 with diffusion data - see supplementary table 2 for demographic information) in order to derive z-scores for each individual and tract. We evaluated the mean differences in normative deviations between patients and controls for each tract using a t-test, applying false discovery rate (FDR) correction (42) to account for multiple comparisons. There were no significant differences in the mean deviations between individuals with psychosis and healthy controls that survived false discovery rate (FDR) multiple comparison correction, although we did find nominally significant effects in the fornix (column and body and the stria terminalis bilaterally). However, we did find evidence for significantly more heterogeneity in individuals with psychosis relative to controls in terms of the proportion of extreme deviations (Z-scores exceeding ±2.6, corresponding to $p < 0.005$). More specifically, individuals with early psychosis had a greater proportion of extreme positive (Mann-Whitney U=1403.0, p=0.0036) and extreme negative (U=1517.0, p=0.0016) Z-scores relative to controls, indicating substantial differences between groups that were highly variable across individuals. Notably, while both positive and negative deviations were present, the frequency of negative outliers (participants with Z-scores < -2.6) was particularly notable in patients relative to controls, highlighting a consistent trend where patients exhibited a greater number of extreme Z-scores across white matter tracts. To assess heterogeneity, we examined the percentage of individuals within each group with such deviations in each tract. Figure 3 visualizes each affected tract, where the colour intensity reflects the percentage of outlier individuals in the group. While both patients and controls exhibit some deviations, patients show a broader spatial distribution across more tracts, whereas controls tend to show deviations in fewer, more circumscribed regions, as illustrated in the bar plots in the supplementary figure 4, with an alternative representation highlighting overlap in individual tracts. These extreme deviations can be interpreted in relation to the underlying microstructural properties of white matter: positive FA deviations ($Z > 2.6$) may reflect greater fibre coherence, increased myelination, or axonal density, whereas negative deviations ($Z < -2.6$) may indicate reduced coherence or disrupted fibre organisation relative to the normative model (44).

We applied msCCA to link FA, cortical thickness, and subcortical volume deviations with PANSS symptom domains (positive, negative, cognitive/disorganised, and total; see Methods and Supplementary text: Multimodal Analysis). The first canonical component showed a significant test correlation ($r = 0.25$, $p = 0.003$; Figure 4A), driven primarily by cognitive and total symptom scores (Figure 4B). Stability selection identified FA in the right uncinate fasciculus and volume loss in the bilateral thalamus as the most informative features (selection probability > 0.8; Figure 4C–D). These deviations were associated with greater symptom severity: higher-than-expected FA in the uncinate and lower-than-expected thalamic volume (Figure 4E–F), suggesting altered fronto-temporal connectivity and subcortical atrophy. Subsequent components showed weaker and less reliable associations ($r = 0.04$, $p = 0.11$; $r = 0.02$, $p = 0.03$), so we focused on the first mode. For details regarding the methodology, results and interpretation please consult the Supplementary text: Multimodal Analysis.

Discussion

This study presents a set of large-scale normative models for FA across major white matter tracts, estimated from a dataset of over 25,000 individuals spanning infancy to old age. Leveraging high-quality multi-shell diffusion MRI data, these models map the trajectory of white matter development and degeneration over the lifespan whilst also quantifying variance across the population. We showcase the clinical utility of these models by mapping inter-individual variation in cohorts of individuals with early psychosis. We show a high degree of inter-individual heterogeneity in these individuals, evidenced by relative increases in both positive- and negative deviations from the normative model in individuals with psychosis relative to controls. These differences were evident despite an absence of case control effects, indicating that the differences were highly individualized. Finally, we show that normative deviations of FA, cortical thickness and subcortical brain volume were accurate multi-modal predictors of symptomatology. Taken together, our findings provide a step toward advancing the understanding of the heterogeneity of white matter alterations in early psychosis.

Our normative models show region-specific developmental trajectories in white matter organization that align well with foundational findings on lifespan changes in FA (15,45). However, we also show that inter-individual variability is considerably higher than the magnitude of lifespan-related changes in FA, underscoring the importance of using approaches such as this to characterize this at the individual level. Studies suggest that increased FA during development relates to synaptic pruning and myelination, while declines in old age are linked to axonal degradation and reduced fiber coherence (46,47). Our models robustly capture these patterns, underscoring their relevance as a normative reference sample and utility for studies examining brain aging and clinical conditions.

In the HCP-EP cohort, we observed high inter-individual variability in white matter organisation in psychosis, consistent with prior findings of structural heterogeneity in both psychosis and other psychiatric conditions (6,7,9,10,48). This variability, seen in the absence of group-level effects, supports the use of normative models for stratifying clinical cohorts (3). Our multivariate results linked symptom severity to decreased thalamic volume and increased FA in the right uncinate fasciculus. These findings align with previous reports implicating the uncinate in psychosis, including reduced FA in recent-onset schizophrenia (14,49,50) and associations with early adversity and emotion recognition deficits (51). The direction of effect in our study differs, potentially reflecting compensatory processes or confounding factors like crossing fibres.

Broader structural studies (15,45,52,53) highlight developmental influences in psychosis, reinforcing the value of normative models for exploring illness trajectories. Longitudinal data will be essential to clarify these effects. See ‘Supplementary text: Interpretation of the results’ for a more detailed interpretation of the results and comparison to existing literature.

One of the benefits of this study is that we focus on acquiring a high-quality diffusion sample with closely harmonized protocols. This maximizes the ability to attribute detected variations to biological differences, rather than artefacts such as data quality or residual site effects. In this study, we focused on FA due to its robustness and directional specificity. While MD is also informative, it is more susceptible to bias from CSF partial volume and partial grey-white mixing. Recent work has shown that such biases can meaningfully alter group differences in ROI-level metrics (54). However, this is only the first step, we intend to augment these models with further models, including other tensor-based metrics, such as mean diffusivity MD and non-tensor models (e.g. neurite orientation dispersion and density

imaging; NODDI (55,56), to take full advantage of the multi-shell diffusion data and provide an even more comprehensive resource for white matter analysis. Finally, we provide these models to the field via our established no-code software platform (23) and via open-source software tools (https://github.com/ramonacirstian/fa_normative_modeling), so that others in the field can easily apply these models to their own data.

We acknowledge some limitations to the current study. our dataset includes fewer participants at the youngest (5–8 years) and oldest (85+ years) age ranges, limiting generalisability at these extremes. While steps were taken to minimise bias, future work should expand coverage as new data become available. Additionally, site effects were addressed, but demographic factors such as socioeconomic background warrant further investigation. We used multiset canonical correlation analysis (msCCA) to link brain deviations to clinical symptoms in a reproducible and interpretable way. Given our moderate sample size, we fixed sparsity parameters to avoid instability from nested cross-validation (57,58). To enhance generalisability and identify robust features, we used permutation testing and stability selection (59). Future work could explore adaptive sparsity tuning in larger samples to improve flexibility.

While harmonisation techniques are commonly used to align datasets across scanners and acquisition protocols, they may suppress variance that is informative in normative modelling. Here, we modelled site effects directly within the statistical framework, enabling site-specific estimation of both means and variances while preserving biologically meaningful variation. This supports generalisable normative modelling, but we acknowledge that some residual scanner effects may remain. Future work may explore combining harmonisation and modelling strategies depending on the intended application. It is important to note that raw FA values were not used in the multivariate analyses due to substantial site-related variability, which we visualised in Figure 1C and Supplementary Figure 6. Normative modelling enabled consistent cross-site comparison by accounting for site scaling differences, thereby offering a more robust basis for linking white matter alterations to symptom severity. These models represent research-ready tools that can support individual-level characterisation of brain deviations and stratification of clinical populations. Their use in everyday clinical practice is still limited by several factors, including the need for calibration data to account for site-related variability, questions about how well current norms apply to more diverse or non-European populations, and the practical challenges of incorporating complex modelling tools into clinical routines. Imaging costs may also be a consideration. Future work is needed to address these challenges and explore how normative models can be made useful in real-world clinical decision-making.

We included self-reported race as a covariate to account for population-level demographic variation, following our prior work (60)**.** We consider this important to reduce the risk of racial bias, but it should be remembered that the datasets on which these models were trained on are not representative of the wider population and are themselves biased towards ‘Western Educated, Industrialised, Rich and Democratic’ (WEIRD) populations (60). Self-reported race is also an imperfect measure, and it is likely that using more flexible modelling approaches may be needed to properly account for these effects (61). For these reasons we also release the models that do not include race so that each researcher using these models can decide for themselves which model is more appropriate for their needs. Future studies should consider incorporating more nuanced demographic information, where available.

In summary, this study provides comprehensive normative reference models for FA across the lifespan, using an extensive dataset that spans infancy to old age. By integrating high-quality diffusion MRI data and using robust modeling techniques, we captured the typical trajectory of white matter development and decline, aligning with prior research and enhancing the field's understanding of brain aging. Our application of these models to a clinical early psychosis cohort underscores their potential utility in identifying atypical white matter patterns in psychiatric conditions. These models not only serve as a benchmark for individual-level assessments but also offer valuable insights for precision medicine, facilitating more personalized interventions. This study highlights the relevance of normative modeling in neuroimaging, paving the way for its integration into clinical and research settings focused on individual variability in brain structure and pathology.

References

1. Borghi E, Onis M, Garza C, Broeck J, Frongillo EA, Grummer-Strawn L. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Stat Med. 2006;25(2):247–65.
2. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. Biol Psychiatry. 2016 Oct;80(7):552–61.
3. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. Mol Psychiatry. 2019 Oct;24(10):1415–24.
4. Rutherford S, Fraza C, Dinga R, Kia SM, Wolfers T, Zabihi M, et al. Charting brain growth and aging at high spatial precision. elife. 2022;11:72904.
5. Bethlehem RAI, Seidlitz J, White SR, Vogel JW, Anderson KM, Adamson C, et al. Brain charts for the human lifespan. Nature. 2022;604(7906):525–33.
6. Wolfers T, Beckmann CF, Hoogman M, Buitelaar JK, Franke B, Marquand AF. Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. Psychol Med. 2020;50(2):314–23.
7. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. Dissecting the heterogeneous cortical anatomy of autism spectrum disorder using normative models. Biol Psychiatry Cogn Neurosci Neuroimaging. 2019;4(6):567–78.
8. Bhome R, Verdi S, Martin SA, Hannaway N, Dobreva I, Oxtoby NP, et al. A neuroimaging measure to capture heterogeneous patterns of atrophy in Parkinson's disease and dementia with Lewy bodies. NeuroImage Clin. 2024;42:103596.
9. Wolfers T, Rokicki J, Alnæs D, Berthet P, Agartz I, Kia SM, et al. Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. Hum Brain Mapp. 2021 Jun;42(8):2546–55.
10. Lv J, Biase M, Cash RF, Cocchi L, Cropley VL, Klauser P, et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. Mol Psychiatry. 2021;26(7):3512–23.
11. Rutherford S, Barkema P, Tso IF, Sripada C, Beckmann CF, Ruhe HG, et al. Evidence for embracing normative modeling. Elife. 2023;12:85082.
12. Remiszewski N, Bryant JE, Rutherford SE, Marquand AF, Nelson E, Askar I, et al. Contrasting case-control and normative reference approaches to capture clinically relevant structural brain abnormalities in patients with first-episode psychosis who are antipsychotic naive. JAMA Psychiatry. 2022;79(11):1133–8.
13. Ge R, Yu Y, Qi YX, Fan Y nan, Chen S, Gao C, et al. Normative modelling of brain morphometry across the lifespan with CentileBrain: algorithm benchmarking and model optimisation. Lancet Digit Health. 2024;6(3):211–21.
14. Sun L, Zhao T, Liang X, Xia M, Li Q, Liao X, et al. Functional connectome through the human life span. BioRxiv. 2024;
15. Lebel C, Beaulieu C. Longitudinal development of human brain wiring continues from childhood into adulthood. J Neurosci. 2011;31(30):10937–47.
16. Kubicki M, Mccarley R, Westin C, Park H, Maier S, Kikinis R, et al. A review of diffusion tensor imaging studies in schizophrenia. J Psychiatr Res. 2007 Jan;41(1–2):15–30.
17. Soares JM, Marques P, Alves V, Sousa N. A hitchhiker's guide to diffusion tensor imaging. Front Neurosci [Internet]. 2013 [cited 2022 Jan 5];7. Available from: http://journal.frontiersin.org/article/10.3389/fnins.2013.00031/abstract
18. Parker CS, Veale T, Bocchetta M, Slattery CF, Malone IB, Thomas DL, et al. Not all voxels are created equal: Reducing estimation bias in regional NODDI metrics using tissue-weighted means. NeuroImage. 2021 Dec;245:118749.
19. Villalón-Reina JE, Moreau CA, Nir TM, Jahanshad N, Individuals Project Consortium SV, Maillard A, et al. Multi-site normative modeling of diffusion tensor imaging metrics using hierarchical Bayesian regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland; 2022. p. 207–17.
20. Zhu AH, Nir TM, Javid S, Villalon-Reina JE, Rodrigue AL, Strike LT. & Alzheimer's Disease Neuroimaging Initiative [Internet]. 2024. Available from: bioRxiv
21. Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, et al. Diffusion tensor imaging: concepts and applications. J Magn Reson Imaging Off J Int Soc Magn Reson Med. 2001;13(4):534–46.
22. Jacobs GR, Coleman MJ, Lewandowski KE, Pasternak O, Cetin-Karayumak S, Mesholam-Gately RI, et al. An introduction to the human connectome project for early psychosis. Schizophr Bull. 2024;sbae123.
23. Barkema P, Rutherford S, Lee HC, Kia SM, Savage H, Beckmann C, et al. Predictive Clinical Neuroscience Portal (PCNportal): instant online access to research-grade normative models for clinical neuroscientists. Wellcome Open Res. 2023;8.

24. Howell BR, Styner MA, Gao W, Yap PT, Wang L, Baluyot K, et al. The UNC/UMN Baby Connectome Project (BCP): An overview of the study design and protocol development. NeuroImage. 2019;185:891–905.
25. Somerville LH, Bookheimer SY, Buckner RL, Burgess GC, Curtiss SW, Dapretto M, et al. The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. Neuroimage. 2018;183:456–68.
26. Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TE, Bucholz R, et al. The Human Connectome Project: a data acquisition perspective. Neuroimage. 2012;62(4):2222–31.
27. Harms MP, Somerville LH, Ances BM, Andersson J, Barch DM, Bastiani M, et al. Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. Neuroimage. 2018;183:972–84.
28. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci. 2016;19(11):1523–36.
29. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage. 2013;80:105–24.
30. S.M. S, F. AA, K.L M. UK Biobank Brain Imaging Documentation version 1.8.
31. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage. 2006;31(4):1487–505.
32. Mori S, Wakana S, Zijl PC, Nagae-Poetscher LM. MRI atlas of human white matter. Elsevier; 2005.
33. Fraza CJ, Dinga R, Beckmann CF, Marquand AF. Warped Bayesian linear regression for normative modelling of big data. NeuroImage. 2021 Dec;245:118715.
34. Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. NeuroImage. 2017 Nov;161:149–70.
35. Nygaard V, Rødland EA, Hovig E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. Biostatistics. 2016 Jan 1;17(1):29–39.
36. Kia SM, Huijsdens H, Rutherford S, De Boer A, Dinga R, Wolfers T, et al. Closing the life-cycle of normative modeling using federated hierarchical Bayesian regression. Yap PT, editor. PLOS ONE. 2022 Dec 8;17(12):e0278776.
37. Bayer JMM, Dinga R, Kia SM, Kottaram AR, Wolfers T, Lv J, et al. Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models. NeuroImage. 2022 Dec;264:119699.
38. Cirstian R, Forde NJ, Andersson JL, Sotiropoulos SN, Beckmann CF, Marquand AF. Objective QC for diffusion MRI data: artefact detection using normative modelling. Imaging Neurosci. 2024;2:1–14.
39. First MB, Williams JB, Karg RS, Spitzer RL. Structured clinical interview for DSM-5 disorders. 2015.
40. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13(2):261–76.
41. Marder SR, Davis JM, Chouinard G. The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials. J Clin Psychiatry. 1997;58(12):538–46.
42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.
43. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. In: The annals of mathematical statistics. 1947. p. 50–60.
44. Beaulieu C. The basis of anisotropic water diffusion in the nervous system – a technical review. NMR Biomed. 2002 Nov;15(7–8):435–55.
45. Westlye LT, Walhovd KB, Dale AM, Bjørnerud A, Due-Tønnessen P, Engvig A, et al. Life-span changes of the human brain white matter: diffusion tensor imaging (DTI) and volumetry. Cereb Cortex. 2010;20(9):2055–68.
46. Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. Longitudinal mapping of cortical thickness and brain growth in normal children. J Neurosci. 2004;24(38):8223–31.
47. Bartók E, Debnár Z, Lázár L. Age-related myelin breakdown: A developmental model of cognitive decline and Alzheimer's disease. Neurobiol Aging. 2004;25(1):5–18.
48. Elad D, Cetin-Karayumak S, Zhang F, Cho KIK, Lyall AE, Seitz-Holland J, et al. Improving the predictive potential of diffusion MRI in schizophrenia using normative models—Towards subject-level classification. Hum Brain Mapp. 2021;42(14):4658–70.
49. McIntosh AM, Maniega SM, Lymer GKS, McKirdy J, Hall J, Sussmann JE, et al. White matter tractography in bipolar disorder and schizophrenia. Biol Psychiatry. 2008;64(12):1088–92.

50. Kawashima T, Nakamura M, Bouix S, Kubicki M, Salisbury DF, Westin CF, et al. Uncinate fasciculus abnormalities in recent onset schizophrenia and affective psychosis: a diffusion tensor imaging study. Schizophr Res. 2009;110(1–3):119–26.
51. Stevens M, Ní Mhurchú S, Corley E, Egan C, Hallahan B, McDonald C, et al. Uncinate fasciculus microstructural organisation and emotion recognition in schizophrenia: controlling for hit rate bias. Front Behav Neurosci. 2024 Mar 19;18:1302916.
52. Si S, Bi A, Yu Z, See C, Kelly S, Ambrogi S, et al. Mapping gray and white matter volume abnormalities in early-onset psychosis: an ENIGMA multicenter voxel-based morphometry study. Mol Psychiatry. 2024 Feb;29(2):496–504.
53. Lebel C, Walker L, Leemans A, Phillips L, Beaulieu C. Microstructural maturation of the human brain from childhood to adulthood. NeuroImage. 2008 Apr;40(3):1044–55.
54. Parker CS, Veale T, Bocchetta M, Slattery CF, Malone IB, Thomas DL. & Alzheimer's Disease Neuroimaging Initiative. Neuroimage. 2021;245:118749.
55. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage. 2012;61(4):1000–16.
56. Kraguljac NV, Guerreri M, Strickland MJ, Zhang H. Neurite orientation dispersion and density imaging in psychiatric disorders: a systematic literature review and a technical note. Biol Psychiatry Glob Open Sci. 2023;3(1):10–21.
57. Ing A, Sämann PG, Chu C, Tay N, Biondo F, Robert G, et al. Identification of neurobehavioural symptom groups based on shared brain mechanisms. Nat Hum Behav. 2019;3(12):1306–18.
58. Savage HS, Mulders PCR, Van Eijndhoven PFP, Van Oort J, Tendolkar I, Vrijsen JN, et al. Dissecting task-based fMRI activity using normative modelling: an application to the Emotional Face Matching Task. Commun Biol. 2024 Jul 20;7(1):888.
59. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Ser B Stat Methodol. 2010;72(4):417–73.
60. Rutherford S, Wolfers T, Fraza C, Harrnet NG, Beckmann CF, Ruhe HG, et al. To which reference class do you belong? Measuring racial fairness of reference classes with normative modeling. 2024.
61. Boer AA, Bayer JM, Kia SM, Rutherford S, Zabihi M, Fraza C, et al. Non-Gaussian normative modelling with hierarchical Bayesian regression. Imaging Neurosci. 2024;2:1–36.

Figure legends

Figure 1. A) Flow chart of the main diffusion image processing steps B) Histogram plot of the data used for normative modeling, showing the population density at each age and highlighting the different datasets used C) Scatterplot exemplifying the quality control process using normative modeling and outlier exclusion based on Z-score thresholding. In this plot, site effects are clearly evident, which are accommodated by the normative models (see Figure 2).

Figure 2. A selection of six white matter tracts and their corresponding normative modelling centile plots highlighting the similarity in white matter formation and degeneration along the lifespan as well as tract specific differences in terms of shapes and variance of the FA values. For visualization purposes, data from different sites are aligned to a common reference (e.g. the mean centiles or the centiles for an arbitrary chosen site) by computing the z-scores separately for each site using the site-specific means and standard deviations, then inverting the z-scores using the mean and standard deviation derived from the common reference. Diffusion imaging derived phenotypes were extracted as mean FA values along the white matter skeleton within each JHU tract. This method, adopted from the UK Biobank processing pipeline, reduces susceptibility to partial volume and misalignment errors. The visual overlays in depict full tract ROIs for spatial reference; analysis was performed using skeletonised values only.

Figure 3: Percentage of individuals with extreme FA deviations ($|Z| > 2.6$) across 48 white matter tracts, shown separately for patients (top) and controls (bottom), and for positive (red) vs. negative (blue) deviations. Each bar represents the percentage of individuals within a group who show an extreme deviation in a given tract. These plots reflect the distribution of deviations across tracts (i.e., the spatial pattern of which tracts are affected and how commonly, within each group). We observe that patients exhibit extreme deviations in a broader range of tracts, whereas controls show a more restricted distribution. This supports the interpretation of increased spatial heterogeneity of white matter alterations in early psychosis. Diffusion imaging derived phenotypes refer to mean FA values extracted from the TBSS skeleton within each tract. The visual overlays depict full tract ROIs for spatial reference; analysis was performed using skeletonised values only.

Figure 4: (A) Density plot of the multiple sparse Canonical Correlation Analysis (msCCA) main components, highlighting the distribution of test canonical correlations separately for each pair of views. (B) Violin plots representing the weights of PANSS symptom scores across the four symptom categories, namely negative symptoms, positive symptoms, cognitive symptoms, and total symptoms. (C) and (D) Selection probabilities for diffusion white matter tracts and Cortical Thickness white matter tracts, respectively, with a red threshold line indicating the chosen selection threshold of 80%. (E) and (F) Glass brain representations of the significantly selected white matter tracts and subcortical regions of interest, respectively. Note that no cortical thickness ROIs survived the selection threshold. The highlighted regions include the uncinate fasciculus (right) for diffusion and the cortical thickness regions: Left-Thalamus and Right-Thalamus

Acknowledgments

Preprint

A preprint of this study is available on BioRxiv

doi: https://doi.org/10.1101/2024.12.11.627897

Funding

This research was supported by grants from the European Research Council (ERC, grant "MENTALPRECISION "10100118) and NIH grant number 1R01MH130362-01A1.

Author contributions:

Conceptualization: RC, NF, AM

Methodology: RC, NF, AM

Validation: RC, NF, AM, NK, GH

Software: RC, NF, AM

Formal analysis: RC, NF, AM

Visualization: RC, NF, AM

Supervision: NF, AM, CB

Writing—original draft: RC

Writing—review & editing: RC, NF, AM, NK, GZ

Financial Disclosure or Competing interests:

CB is director and shareholder for SBGneuro

All other authors report no biomedical financial interest or potential conflicts of interest.

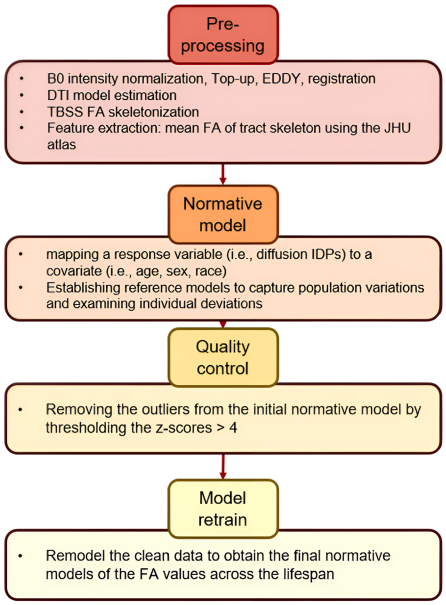
Data and materials availability:

The data used in the present study is part of the UK Biobank dataset which is available to be downloaded upon completing an access application. More information can be found on the dedicated webpage (UK Biobank, n.d.). The code used to process the data and train the normative models is also available online on GitHub (https://github.com/ramonacirstian/fa_normative_modeling n.d.)

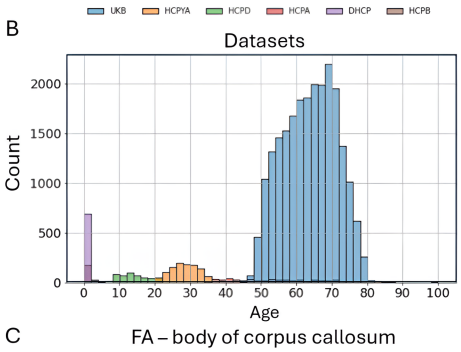
Supplement Description:

Supplement Methods, Discussion, Figures S1-S6, Tables S1-S3

A



B



C