Probing the dysmyelination hypothesis in schizophrenia spectrum and bipolar disorders

Methodological appraisal and clinical investigations of cerebral grey and white matter myelination

Doctoral thesis
Stener Nerland
Submitted for evaluation 15 May, 2023

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0281-2

Cover: UiO.
Print production: Graphic center, University of Oslo.

Table of contents

Study 1-3

I List of publications

Study 1

Multisite reproducibility and test-retest reliability of the T1w/T2w-ratio: A comparison of processing methods

Nerland S, Jørgensen KN, Nordhøy W, Maximov II, Bugge RAB, Westlye LT, Andreassen OA, Geier OM, Agartz I. (2021). Neuroimage. 2021;245:118709.

Study 2

Assessing regional intracortical myelination in schizophrenia and bipolar spectrum disorders using the optimized T1w/T2w-ratio

Jørgensen KN and Nerland S, Slapø NB, Norbom LB, Mørch-Johnsen L, Wortinger LA, Barth C, Dimitrios A, Maximov II, Geier OM, Andreassen OA, Jönsson EG, Agartz I. In preparation.

Study 3

Widespread alterations of diffusion tensor imaging metrics in patients with schizophrenia without current auditory hallucinations

Nerland S, Slapø NB, Barth C, Mørch-Johnsen L, Jørgensen KN, Beck D, Wortinger LA, Westlye LT, Jönsson EG, Andreassen OA, Maximov II, Geier OM, Agartz I. Under review.

II Summary

Introduction

Schizophrenia spectrum and bipolar disorders are severe mental disorders that can significantly affect an individual's experience of the world and quality of life. The pathophysiology and aetiology of these disorders are largely unknown, but dysmyelination has been proposed as a neurobiological mechanism. It has been suggested that abnormal myelination disrupts basic sensory processing, leading to auditory hallucinations and delusions of control. Magnetic resonance imaging (MRI) is a uniquely flexible tool for assessing myelin-related tissue properties. In this doctoral project, we employed two MRI-based neuroimaging techniques sensitive to myelin in the cerebral cortex and white matter. The ratio of T1- and T2-weighted MRI images (T1w/T2w-ratio) and diffusion tensor imaging (DTI) metrics.

Aims and objectives

The aims of this doctoral project were first to critically assess the T1w/T2w-ratio as a measure of intracortical myelin and second to investigate putative alterations of myelin in patients with schizophrenia spectrum and bipolar disorders. To achieve these aims, we compared T1w/T2w-ratio maps and DTI metrics in patients with those of healthy controls and explored associations with age, antipsychotic medication, and psychotic symptoms. The first study aimed to assess the measurement properties of the T1w/T2w-ratio and gauge the effects of corrections for partial volume effects, surface outliers, and intensity nonuniformities, as well as intensity normalisation. In the second study, we used an optimised T1w/T2w-ratio pipeline from the first study to compare intracortical myelin in patients with schizophrenia spectrum and bipolar disorders with that of healthy controls. Finally, in the third study, we aimed to explore the relationship between DTI metrics and auditory hallucination status in patients with schizophrenia spectrum disorders.

Materials and methods

In the first study, we designed and implemented an in-house test-retest study and included two additional openly available datasets. We created optimised pipelines for computing the T1w/T2w-ratio and compared their test-retest reliability and agreement with expected myeloarchitecture. In the second study, we included patients with schizophrenia spectrum (n=64) and bipolar disorders (n=91) and compared their cortical T1w/T2w-ratio maps, computed with an optimised pipeline from the first study, with those of healthy controls (n=155). In the third study, we performed diffusion tractography in patients with schizophrenia spectrum disorders (n=140) and healthy controls (n=140) and compared DTI metrics between patients, with and without current and lifetime auditory hallucinations, and healthy controls.

Results

We observed low test-retest reliability of the raw T1w/T2w-ratio and, for some datasets, poor agreement with known myeloarchitecture. These issues could be improved with some, but not all, intensity normalisation and intensity nonuniformity correction methods. In the second study, T1w/T2w-ratio maps did not differ significantly between patients with schizophrenia spectrum disorders or bipolar disorders relative to healthy controls. However, significant interactions with age indicated more positive age-related T1w/T2w-ratio trajectories in patients with schizophrenia spectrum disorders. This effect was associated with both antipsychotic medication use and dose. In the third study, we found higher mean diffusivity (MD) and radial diffusivity (RD) across widespread fibre tracts in patients with schizophrenia spectrum disorders without current auditory hallucinations compared to healthy controls. In contrast, patients with current auditory hallucinations only showed limited alterations of DTI metrics relative to healthy controls. Dimensionality reduction enhanced sensitivity to group-level differences in axial diffusivity (AD).

Conclusions

Intensity normalisation and intensity nonuniformity correction improved test-retest reliability and agreement with expected myeloarchitecture of the T1w/T2w-ratio, particularly in the presence of scanner-related intensity nonuniformities. The divergent T1w/T2w-ratio age trajectories may be consistent with a hypothesised promyelinating effect of antipsychotic medication. The limited relationship with current auditory hallucinations calls into question the notion that altered DTI metrics reflect specific abnormalities underlying auditory hallucinations. Conversely, higher MD and RD in patients without auditory hallucinations suggest myelin-related alterations in this group. The results of this doctoral project are consistent with an involvement of myelin in patients with schizophrenia spectrum disorders and emphasise the importance of assessing the validity of myelin imaging techniques.

III Summary in Norwegian

Bakgrunn

Schizofrenispektrumlidingar og bipolare lidingar er alvorlege mentale lidingar som kan påverke korleis ein oppfattar verkelegheita og føre til redusert livskvalitet. Patofysiologien og etiologien til desse lidingane er uviss men dysmyelinering er utpeikt som ein mogleg nevrobiologisk mekanisme. Det har blitt foreslått at avvikande myelinisering kan forstyrre grunnleggande sanseprosessering og føre til hørselshallusinasjonar. Magnetresonanstomografi (MR) er ein særs fleksibelt reiskap for å vurdere eigenskapar ved vevet som er relaterte til myelin. I dette doktorgradsprosjektet fokuserte vi på to MR-baserte teknikkar som er sensitive til myelin-relaterte forandringar i hjernebarken og kvit substans. Ratioen av T1- og T2-vekta MR bilete (T1w/T2w-ratioen) og diffusjon tensor avbilding (DTI).

Mål

Måla med dette doktorgradsprosjektet var først å gjere ei kritisk vurdering av T1w/T2w-ratioen som eit mål på myelin i hjernebarken, og so deretter undersøke moglege endringar i myelin hjå pasientar med schizofrenispektrumlidingar og bipolare lidingar. For å nå desse måla samanlikna vi T1w/T2w-ratioen og DTI-mål hjå pasientar og friske kontrolldeltakarar og vi utforska samanhengane mellom desse måla og alder, bruk av antipsykotiske medisinar og psykotiske symptom. I den første studien var målet å evaluere måleeigenskapane til T1w/T2w-ratioen og vurdere effekten av ulike korreksjonsmetoder og intensitetsnormalisering. I den andre studien brukte vi ei optimalisert utrekning av T1w/T2w-ratioen for å jamføre myelin i hjernebarken hjå pasientar med schizofrenispektrumlidingar og bipolare lidingar med friske kontrollar. I den tredje studien var målet å utforske samanhengar mellom DTI-mål og hørselshallusinasjonar hjå pasientar med schizofrenispektrumlidingar.

Metoder

Først gjennomførte vi ein test-retest studie og undersøkte i tillegg to fritt tilgjengelege datasett. Vi laga optimaliserte prosesseringsstraumar for å rekne ut T1w/T2w-ratioen og vi vurderte test-retest reliabiliteten og samsvaret med kjend myeloarkitektur. I den neste studien inkluderte vi pasientar med schizofrenispektrumlidingar (n=64) og bipolare lidingar (n=91) og samanlikna T1w/T2w-ratio kart i hjernebarken mellom desse gruppene og friske kontrolldeltakarar (n=155). I den tredje og siste studien utførte vi diffusjonstraktografi hjå pasientar med schizofrenispektrumlidingar (n=140) og friske kontrolldeltakarar (n=140), og samanlikna DTI-mål mellom pasientar, med og utan hørselshallusinasjonar, med friske kontrollar.

Resultat

Vi fann låg test-retest reliabilitet for den ikkje-optimaliserte T1w/T2w-ratioen, og for nokre datasett var samsvaret med kjend myeloarkitektur dårleg. Det var mogleg å forbetre dette med nokre av metodene vi testa for å korrigere intensitet og ujamnheiter, medan andre metodar ikkje gav gode resultat. I den andre studien fann vi ingen forskjell mellom pasientar og kontrollar i verdiane til T1w/T2w-ratioen. Vi såg derimot statistisk signifikante interaksjonar med alder. Desse tyda på meir positive samanhenger mellom alder og T1w/T2w-ratioen hjå pasientar med schizofrenispektrumlidingar samanlikna med dei friske kontrolldeltakarane. Denne effekten hadde samanheng med både bruk og dose av antipsykotiske medisinar. I den tredje studien fann vi høgare gjennomsnittsdiffusivitet (MD) og radial diffusivitet (RD) på tvers av mange fibertrakter hjå pasientar med schizofrenispektrumlidingar utan aktuelle hørselshallusinasjonar samanlikna med friske kontrolldeltakarar. Vi såg derimot berre avgrensa forskjellar i DTI-mål mellom pasientane med aktuelle hørselshallusinar og friske kontrolldeltakarar. Når vi undersøkte variasjon på tvers av fibertrakter auka sensitiviteten til gruppeforskjellar i aksial diffusivitet (AD).

Konklusjonar

Normalisering av signalintensitet og korreksjon for intensitetsujamnheiter gav høgare test-retest reliabilitet og betre samsvar med kjend myeloarkitektur. Vi såg særskild forbetring der det frå før var store intensitetsujamnheiter relaterte til MR-skannaren. Samanhengen mellom T1w/T2w-ratioen og alder var annleis hjå pasientar med schizofrenispektrumlidingar og resultata tyda på ein samanheng med bruk av antipsykotiske medisinar, med ein dose-response samanheng. Dette kan vere konsistent med hypotesa om at antipsykotiske medisinar har ein promyelinerande effekt. Den avgrensa samanhengen med aktuelle hørselshallusinasjonar gjev grunn til å sette spørsmålsteikn ved tanken om at endringar i DTI-mål reflekterer spesifikke avvik som ligg bak desse opplevelsane. Høgare MD og RD hjå pasientane som ikkje hadde aktuelle hørselshallusinasjonar tyder på endringar relatert til myelin i denne gruppa. Resultata av dette doktorgradsprosjektet er i tråd med hypotesa om at myelin er involvert i patofysiologien til schizofrenispektrumlidingar og syner kor viktig det er å vurdere validiteten til metoder for å avbilde myelin.

IV Acknowledgements

I would first like to thank my main supervisor Ingrid Agartz for her guidance and support over the years and for always encouraging me to pursue exciting research questions. She is a true scientist who never stops seeking answers and has been a major source of inspiration for me. I am also grateful to my co-supervisors, Oliver Geier and Ivan Maximov. Thank you for your patience and insight and always being available to answer my questions, no matter how misguided.

This project was possible only with the help and support of numerous colleagues and collaborators. First, I wish to thank my colleagues in the Imaging Psychosis group, Kjetil Jørgensen, Claudia Barth, Lynn Mørch-Johnsen, Laura Wortinger, Dimitrios Andreou, Runar Smelror, Anita Banafsheh, and Arielle Crestol. I also want to acknowledge former colleagues, whom I will refrain from naming to avoid any embarrassing omissions. My time here would not have been the same without you.

I would also like to thank my coauthors for sharing their expertise and knowledge and for their help in ironing out the many kinks and wrinkles that inevitably arise. These collaborations have been the highlight of my time as a doctoral student. Wibeke Nordhøy, in particular, played an instrumental role in the first publication of this doctoral project, and her expertise was critical to the success of this project.

A special mention goes to Kjetil Jørgensen, who has not only been a close collaborator but also a dear friend. I truly appreciate his curiosity, sense of humour, and thirst for knowledge, which have constantly motivated and encouraged me. Our lively discussions about everything from philosophy to phrenology have been incredibly enjoyable, and I look forward to many more in the future!

I want to thank the South-Eastern Norway Regional Health Authority for funding my doctoral project. I also want to thank Kari Agnes and Ragne Gjengedal at the Unit for Clinical Research and Innovation at Diakonhjemmet Hospital. Their support before and during this project has been invaluable, and I am deeply thankful for their efforts in ensuring the smooth operation of all matters practical, whether big or small.

Finally, I would like to express gratitude to my wife, Hsiao Yu, for her support throughout my academic journey. Thank you for putting up with me when there is "just one more analysis" to be run and for always helping me get my priorities in order. Your intelligence, thoughtfulness, and companionship have been an incalculable source of happiness to me.

V Abbreviations

MRI	Magnetic resonance imaging
DTI	Diffusion tensor imaging
RF	Radiofrequency
TR	Repetition time
TE	Echo time
FA	Fractional anisotropy
MD	Mean diffusivity
RD	Radial diffusivity
AD	Axial diffusivity
qMRI	Quantitative magnetic resonance imaging
MWF	Myelin water fraction
MT	Magnetisation transfer
MTR	Magnetisation transfer ratio
DKI	Diffusion kurtosis imaging
NODDI	Neurite orientation dispersion and density imaging
ENIGMA	Enhancing Neuroimaging Genetics through Meta-Analysis
GWC	Grey-white contrast
LAPC	Language and auditory processing circuitry
AF	Arcuate fasciculus
NORMENT	Norwegian Centre for Mental Disorders Research
OUS	Oslo University Hospital
HUBIN	Human Brain Informatics
TOP	Thematically Organised Psychosis
HCP	Human Connectome Project

DSM	Diagnostic and Statistical Manual of Mental Disorders
SCID	Structured Clinical Interview
CPZ	Chlorpromazine-equivalent dose
WASI	Wechsler Abbreviated Scale of Intelligence
WAIS	Wechsler Adult Intelligence Scale
GAF	Global Assessment of Functioning Scale
AUDIT	Alcohol Use Disorder Identification Test
DUDIT	Drug Use Disorders Identification Test
PANSS	Positive and Negative Syndrome Scale
SCAN	Schedules for Clinical Assessment in Neuropsychiatry
SAPS	Scales for the Assessment of Positive Symptoms
SANS	Scales for the Assessment of Negative Symptoms
ICV	Intracranial volume
HCP-MPP	Human Connectome Project - Minimal Processing Pipelines
FSL	FMRIB Software Library
ROI	Region of interest
SAMSEG	Sequence Adaptive Multimodal Segmentation
HCP-YA	Human Connectome Project - Young Adult
HCP-MMP	Human Connectome Project - Multi-Modal Parcellation
CV	Coefficient of variation
ICC	Intraclass correlation coefficient
PCA	Principal component analysis
B1+	B1 transmit
B1-	B1 receive
NAWM	Normal-appearing white matter
TBSS	Tract-based spatial statistics

1. Introduction

Schizophrenia spectrum and bipolar disorders are severe mental disorders that affect critical cognitive functions such as language, thought, and perception. They can have a significant detrimental impact on quality of life and psychosocial functioning. While their pathophysiology and aetiology remain elusive, dysmyelination has been proposed as an underlying biological mechanism, with supporting evidence from genetic, neuroimaging, and post-mortem studies.

Magnetic resonance imaging (MRI) is a uniquely flexible tool that is widely used to study brain structure and function in severe mental disorders. Despite its functional importance, direct in vivo imaging of myelin is challenging. Several MRI-based proxy measures of cerebral myelination have been proposed, but more knowledge about their measurement properties is needed. In this doctoral project, we investigated methodological issues related to in vivo imaging of intracortical myelin as well as clinical research questions related to myelin in grey and white matter.

In the introduction, we give an overview of the clinical characteristics of schizophrenia spectrum and bipolar disorders with a particular focus on auditory hallucinations and antipsychotic medication. We present the general principles of MRI, describe the MRI pulse sequences employed in this project, and discuss some challenges related to assessing myelin with MRI. Finally, we present some of the previous research on myelin-related alterations in psychotic disorders.

1.1 Schizophrenia spectrum and bipolar disorders

Schizophrenia spectrum and bipolar disorders are leading causes of disability worldwide.[1] Psychotic symptoms are commonly classified as 'positive', i.e., the presence of unusual experiences or behaviour, or 'negative', i.e., a lack of typical experiences or behaviour. Positive symptoms include delusions, disorganised speech, and hallucinations, i.e., sensory percepts in the absence of corresponding sensory stimuli. Negative symptoms include avolition, apathy, alogia (poverty of speech), social withdrawal, and reduced emotional affect. There is substantial heterogeneity in the clinical presentation, and the symptomatology of the disorders includes cognitive, motor, and mood symptoms in addition to positive and negative symptoms.[2] Psychotic episodes are associated with cognitive deficits, sleep disturbances, anxiety, and depression. Furthermore, many patients experience co-occurring psychiatric and somatic health issues (i.e., comorbidity) in addition to the core symptoms of their diagnoses.

Individuals diagnosed with schizophrenia spectrum or bipolar disorders often have impaired occupational and psychosocial function. The disorders are associated with elevated mortality[3,4] and can come at a great cost to affected individuals, caregivers, and society at large. An organic neurological basis of psychosis has long been suspected, and the notion that it is caused by abnormal brain connectivity can be traced back to the 19th century.[5] Notably in the form of the *Sejunktionshypothese* of Carl Wernicke (1848-1905), which proposed that disrupted connectivity within association fibres leads to symptoms such as hallucinations.[5–7] Despite considerable efforts, the underlying pathophysiological mechanisms of schizophrenia spectrum and bipolar disorders remain largely unknown.

Schizophrenia spectrum and bipolar disorders have been proposed to exist along a psychosis spectrum,[8–10] in part due to substantial genetic overlap and psychotic symptoms being common in bipolar disorders.[11–13] The lifetime prevalence of psychotic symptoms has been estimated to exceed 60% in bipolar I disorder[14,15] but varies between studies.[16] We will occasionally use the term "psychotic disorders" to refer to schizophrenia spectrum disorders and bipolar disorders with psychotic features but note that psychotic symptoms are not confined to these groups and that not all individuals diagnosed with a bipolar disorder experience psychotic symptoms.

1.1.1 Schizophrenia spectrum disorders

Schizophrenia spectrum disorders have a lifetime prevalence of over 1%[17] and are considered to be some of the most severe psychiatric disorders. Important diagnostic criteria include delusions, hallucinations, disorganised speech (e.g., frequent derailment or incoherence), grossly disorganised or catatonic behaviour, and negative symptoms. Schizophrenia spectrum disorders are often chronic illnesses with incomplete remissions and frequent relapses, and symptom remission does not necessarily imply functional recovery.[18,19] Schizophrenia spectrum disorders are associated with lower rates of employment[20] and impaired quality of life.[21]

Positive symptoms tend to be the most conspicuous and often have their onset in adolescence or in early adulthood. The acute phase of the illness, characterised by positive symptoms, is often preceded by a so-called prodromal phase, in which social withdrawal, thought disturbance and negative symptoms may be present. Comorbidities such as substance abuse, anxiety disorders, and cardiovascular disease are common. There is an ongoing debate whether schizophrenia spectrum disorders represent one or several distinct illnesses.[22] This has led researchers to recommend a dimensional approach as a way to disentangle the clinical heterogeneity of schizophrenia spectrum disorders,[23] and has intensified the search for biological markers to identify relevant subgroups.

1.1.2 Bipolar disorders

Bipolar disorders, including bipolar I and bipolar II disorders, have an estimated lifetime prevalence of around 2.4%, with a high variability in estimates.[24] Bipolar disorders are characterised by mood swings in the form of periods of major depression punctuated by, often briefer, periods of mania or hypomania. These are characterised by increased energy or activity, which may be recognised as elevated, expansive or markedly irritable mood. Manic episodes are associated with inflated self-esteem or grandiosity, decreased need for sleep, flight of ideas or racing thoughts, distractibility, increase in goal-directed behaviour, and hedonistic behaviour with the potential for adverse consequences. The diagnostic criteria for bipolar II disorder are similar to those of bipolar I disorder but do not include a history of mania. Instead, hypomanic episodes are required, with an observable and marked change from the usual non-depressed mood. Manic episodes are required to have a duration of at least seven days or require hospitalisation, whereas hypomanic episodes are only required to have a duration of at least four days. Psychotic symptoms are common in bipolar I disorder, especially during manic episodes, but they can also occur during depressive episodes.[25] They are, by definition, absent in hypomanic episodes and occur mainly during depressive episodes in bipolar II disorder. Treatment of bipolar disorders includes mood stabilisers, e.g., lithium and antiepileptic medication, and antipsychotic medication.

1.1.3 Auditory hallucinations

Auditory hallucinations, defined as auditory percepts without an external source, are considered to be core symptoms of schizophrenia spectrum disorders, where they have a lifetime prevalence of 60-80%.[26] They are also frequently reported in bipolar disorders, especially during manic episodes.[27] While hallucinations in other sensory modalities, e.g., visual, somatic/tactile, and olfactory hallucinations, are common in schizophrenia spectrum disorders,[28] auditory hallucinations are the most frequent. A majority of those who report hallucinations in other modalities also report auditory hallucinations.[28] Auditory hallucinations most often present as voices, but nonverbal hallucinations, such as music, animal sounds, clicks, and ringing, are also common.[29,30] A notable feature of auditory verbal hallucinations in psychotic disorders is the negative valence of the voices. A large phenomenological survey of patients with psychotic disorders found that the predominant voice is more often described as angry, authoritative, malicious or bossy than as loving, gentle, kind or friendly.[30] Auditory hallucinations are not specific to psychotic disorders and can also occur in other somatic and psychiatric disorders and in non-clinical populations.[31,32] However, phenomenological differences have been reported between hallucinations in psychotic disorders and those experienced in other conditions. For instance, negative voice content is more common in clinical compared to non-clinical hallucinators[33] and is linked to greater experienced distress and lower functioning.[34]

The pathophysiology of auditory hallucinations remains largely unknown, but disrupted source monitoring has been advanced as a possible explanation.[35,36] In this model, internal events, e.g., thoughts or inner speech, are misidentified as externally generated, leading to hallucinations.[37,38] It has been framed in the context of forward modelling, the mechanism by which the sensory consequences of self-initiated actions are predicted and attenuated.[39–44] Within this framework, prediction is accomplished through efference copies, i.e., copies of the motor signals that elicit corollary discharges that, in turn, suppress activation of sensory regions. This mechanism has been studied in both humans and non-human animals.[45] In particular, excitatory neurons in the auditory cortex have been shown to be inhibited by local parvalbumin-positive interneurons in mice.[46,47]

Disruption of the efference copy/corollary discharge mechanism has been proposed to underlie auditory hallucinations and delusions of control in psychotic disorders.[39] Importantly, this mechanism can be disrupted at multiple levels. First, neural assemblies initiating actions may fail to issue the appropriate efference copies. Second, corollary discharges may fail to suppress activation within the sensory regions, e.g., due to dysfunction of parvalbumin-positive interneurons.[48] Third, the corollary discharge may not be correctly transmitted by the issuing regions, e.g., due to disrupted neural synchrony caused by conduction delays. In section 1.1.6, we discuss abnormal myelination in psychotic disorders and how it may relate to this model of auditory hallucinations. See **Figure 1** for an illustration of the efference copy/corollary charge mechanism in the context of self-generated speech.

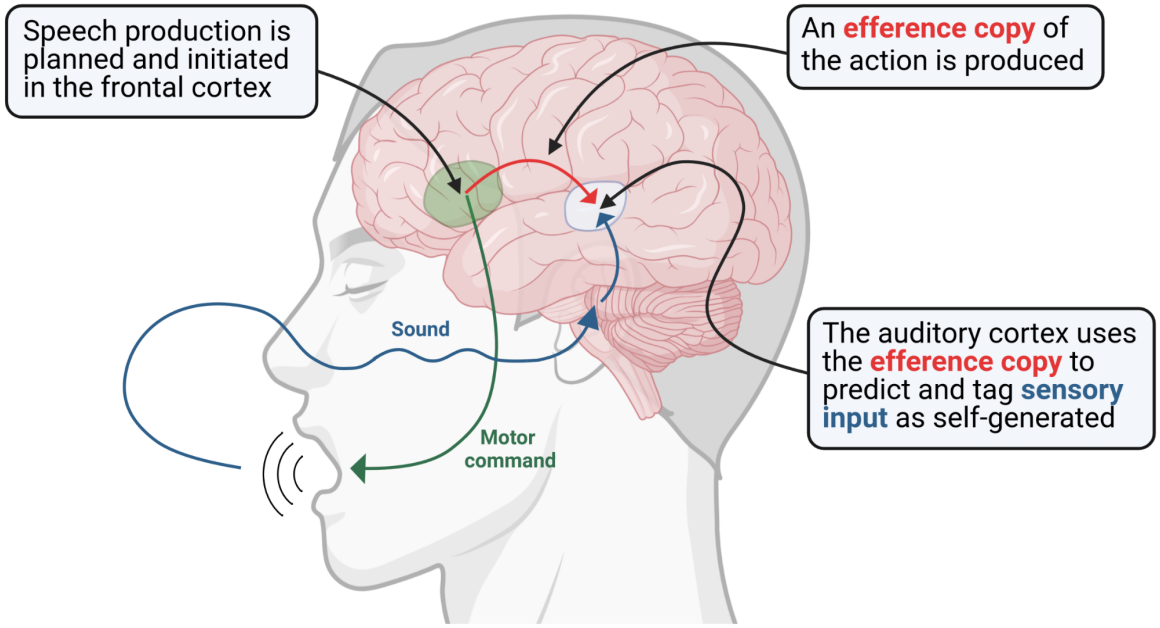


Figure 1. Simplified schematic of forward modelling in the context of speech. The action plan issues from the frontal cortex (green circle near Broca's area), creating a motor command (green arrow) and an efference copy (red arrow). The efference copy generates a corollary discharge in the auditory cortex (blue circle) used to predict sensory input from the self-generated sound (blue arrows). As a result, neural activation is attenuated, and the sensory input is tagged as self-generated.

1.1.4 Antipsychotic medication

The most effective medical intervention for psychotic disorders currently available is treatment with antipsychotic medications.[49] Although there are substantial differences between the receptor binding profiles of different types of antipsychotic medications, they share the pharmacological mechanism of D_2 dopamine receptor antagonism.[50] Discovery of this mechanism led to the formulation of the dopamine hypothesis of psychosis in the 1960s, and it remains one of the most influential and longest-standing models of psychosis. In its original formulation, it states that positive symptoms arise as a consequence of hyperactive dopamine transmission in the mesolimbic pathway, but it was later revised to also accommodate for hypoactive dopamine transmission in the prefrontal cortex. As new neuroscientific evidence is uncovered, the dopamine hypothesis has been frequently updated and refined.[51]

While the evidence for the dopamine hypothesis is compelling, dopamine receptor binding alone does not guarantee the absence of psychotic symptoms. Many patients experience persisting symptoms, and relapses are common.[52] These considerations, along with studies showing widespread anomalies in brain structure and function, have led to a renewed interest in complementary or alternative models of psychosis. It has been suggested that dopamine dysfunction represents a "final common pathway" and that different brain abnormalities may lead to similar dopamine signalling abnormalities.[53] Such complementary models of psychosis include disrupted functional and structural connectivity,[54] the involvement of other neurotransmitters such as glutamate,[55] neurodevelopmental disturbances,[56] and, as we will discuss in the next two sections, abnormal myelination.

1.1.5 The role of myelin in the brain

Myelin is present in both grey and white matter and is critical for the function of the central and peripheral nervous systems. It consists of a lipid-rich substance that forms an insulating sheath around the axons of neurons, where it adjusts the speed of neural conductance. The transmission of action potentials is 10-100 times faster for myelinated axons than for unmyelinated axons due to the mechanism of saltatory conduction. Importantly, myelination is not just a question of axons being myelinated or unmyelinated but represents a dynamic process that continues into adulthood. Continuous changes in myelination fine-tune neural activity in short- and long-range neural circuitry,[57] which is essential for higher-order functions dependent on neuroplasticity, such as learning and memory.[58] Myelin in the central nervous system is formed and maintained by oligodendrocytes, non-neuronal support cells, in a process that can be modulated by other glial cells, particularly astrocytes.[59] In addition to increasing and regulating conduction velocities, myelin and myelinating oligodendrocytes facilitate axonal transport and influence axonal energy metabolism and therefore serve an important metabolic role in the brain.[60]

Myelin is abundant in white matter, where it envelops the axons of intrahemispheric cortico-cortical fibre tracts, interhemispheric fibre tracts, and fibre tracts projecting to areas such as deep nuclei, the brainstem, cerebellum, or the spine. These fibre tracts are commonly known as association, commissural, and projection tracts, respectively. In grey matter, myelin is less dense and is mainly located on the axons of short-range inhibitory interneurons, where it plays a crucial role in regulating the excitation/inhibition balance.[61] Myelination of both grey and white matter continues through adolescence and into early adulthood with regional differences in the rates of myelin maturation and distinct waves of growth.[62,63]

Studies have linked myelin to cognitive functions such as error processing, processing speed,[64] and cognitive control[65] in healthy adults. Additionally, myelination within the first year of life has been associated with language development.[66] The distribution of myelin in the cerebral cortex, known as *myeloarchitecture*, is not uniform and varies across cortical laminae and regions. This provides an alternative perspective on brain structure compared to *cytoarchitecture*, which is based on the characterisation of cell types and densities. The most comprehensive myeloarchitectonic studies were conducted between 1910 and 1970 by Cécile Vogt-Mugnier and Oskar Vogt and their collaborators.[67] See **Figure 2** for a comparison of maps depicting cyto- and myeloarchitecture.

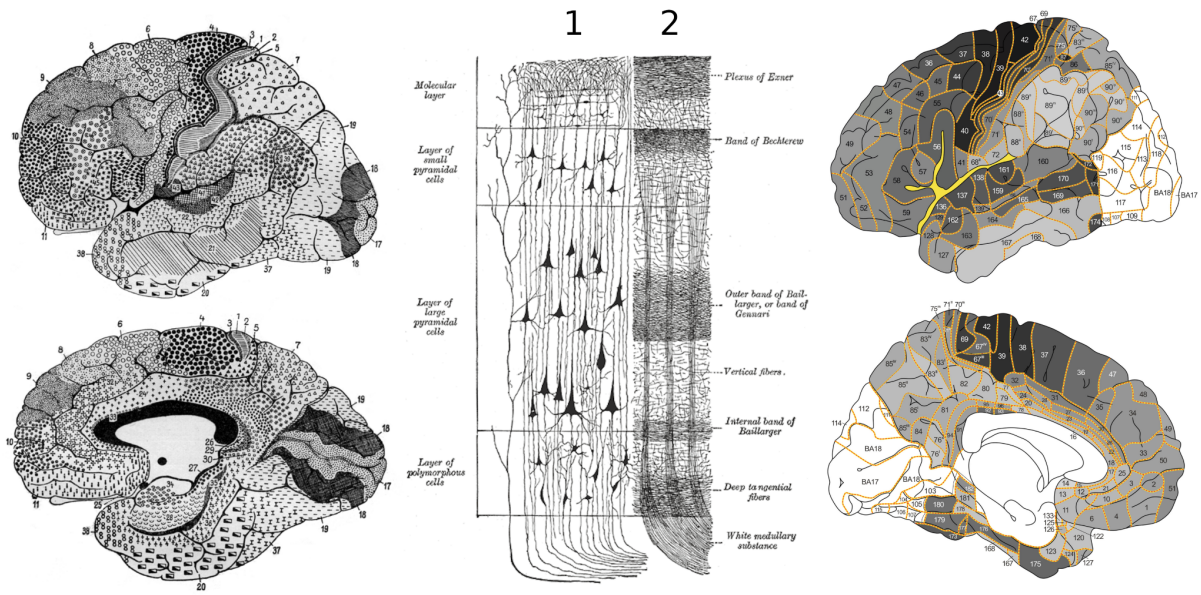


Figure 2. Left: Cytoarchitectonic map by Korbinian Brodmann and adapted from Nieuwenhuys and Broere.[67] Middle: Cortical layers in the cerebral cortex with 1. cell distributions, and 2. myeloarchitecture. Illustration by Henry Gray.[68] Right: Myeloarchitectonic map based on myelin-staining studies from the Vogt-Vogt school, aggregated and aligned to a standard brain by Nieuwenhuys and Broere.[69] Darker areas indicate more dense myelination. Note that the occipital lobe is highly myelinated but appears white due to a lack of studies on this region.

1.1.6 The dysmyelination hypothesis

Abnormal myelination has been implicated in psychotic disorders.[70] Some genetic studies, but not all,[71] have found downregulation of genes related to oligodendrocytes and myelination,[72–74] and genetic analyses have implicated a glia-oligodendrocyte pathway in schizophrenia spectrum disorders.[75] Histological studies in both schizophrenia spectrum and bipolar disorders have shown reduced numerical density, abnormal morphology of oligodendrocytes,[76–78] and lower myelin staining intensity.[79,80] Finally, neuroimaging studies report alterations consistent with abnormal myelination of both grey and white matter in psychotic disorders. These findings are further discussed in section 1.3. More indirectly, myelination continues into adolescence and early adulthood, coinciding with an elevated risk of developing a psychotic disorder.[62,81] Taken together, the evidence suggests that myelination is affected in psychotic disorders. Importantly, the evidence comes from several levels of observation, i.e., genetic associations, ultrastructural abnormalities of myelinating glial cells, and in vivo neuroimaging studies.

Whitford et al[82] proposed that abnormal myelination of frontally projecting white matter fibre tracts is a causal mechanism in schizophrenia spectrum disorders. Specifically, they conjectured that dysmyelination results in conduction delays in the efference copies generated by volitional actions, resulting in delayed corollary discharges that fail to adequately suppress the sensory consequences of those actions. An attractive feature of this hypothesis is that it can also account for the dopaminergic hyperactivity associated with psychotic symptoms, which is considered a consequence of dysmyelination-induced conduction delays. However, this may only apply to a subset of patients, and other brain abnormalities may lead to the same dopaminergic dysfunction in line with the "final common pathway" hypothesis.

Although much research on myelin-related alterations in psychotic disorders has focused on white matter, the cerebral cortex is also myelinated and shows a similar region-specific maturational trajectory as frontally projecting white matter fibre tracts. Notably, Bartzokis and colleagues have performed several studies testing the hypothesis that abnormal intracortical myelination is central to the pathophysiology of schizophrenia spectrum disorders.[83–87] These studies are discussed in section 1.3.4.

Studies have linked serum lipid levels to the clinical efficacy of antipsychotic medication,[88–90] providing indirect support for the dysmyelination hypothesis. It has been proposed that the therapeutic action of antipsychotics may be related to promyelinating effects.[83,91] Interestingly, preclinical studies have reported lipogenic effects of many psychotropic drugs[92,93] as well as ameliorative effects of antipsychotics on cuprizone-induced demyelination deficits.[94–96] If confirmed, such a link would deepen our understanding of the clinical response to antipsychotics and could provide targets for the development of new medical interventions.[91,97]

1.2 Magnetic resonance imaging

The use of MRI has revolutionised the study of a wide range of somatic and psychiatric conditions. It relies on the nuclear magnetic resonance effect in which nuclei subjected to a strong magnetic field are perturbed by a weaker oscillating magnetic field, causing the emission of an electromagnetic signal. Compared to other medical imaging techniques, e.g., computed tomography, MRI offers superior contrast in soft tissues such as the brain. By adjusting the parameters of MRI pulse sequences, a great variety of signal contrasts can be generated, enabling the measurement of a wide variety of properties of brain structure, function, and metabolism. Importantly, it is a safe, non-invasive procedure that does not employ ionising radiation and can be applied to the study of a wide range of conditions with minimal risk. In this section, we present the general principles of MRI and describe the MRI pulse sequences that were used in this project.

1.2.1 General principles

Biological tissue typically consists of 60-80% water in which macromolecules are suspended. A major constituent of water is the hydrogen isotope protium, which consists of one proton, with a spin of ½, and one electron. Nuclear spins are associated with a magnetic moment, which is small for individual protons but can be summed across many protons to yield the macroscopic magnetisation of the sample. This net magnetisation is proportional to the strength of the magnetic field experienced by the protons, which necessitates the use of strong magnets to induce a measurable macroscopic magnetisation. Since protons have a spin of ½, their spins are limited to two spin eigenstates corresponding to low and high energy, spin down (-½) and spin up (+½) respectively. The distribution of spins in each state can be computed using Boltzmann statistics, which shows that only about 1 proton out of every 50,000 protons contributes to the net magnetisation when subjected to a magnetic field of 3 Tesla at room temperature. Nevertheless, since 1 mm^3 of water contains about 6.68×10^{18} protons, this effect becomes large at greater scales.

At equilibrium, the macroscopic magnetisation vector, M, points in the direction of the main magnetic field, conventionally denoted as the z-axis. By applying a second oscillating magnetic field, known as a radiofrequency (RF) pulse, the magnetisation vector can be rotated away from the z-direction and towards the xy plane. As M returns to equilibrium, it precesses around the direction of the magnetic field at an angular frequency known as the Larmor frequency, which depends on the strength of the magnetic field and the tissue being imaged. The oscillation of the RF pulse is usually chosen to be at the Larmor frequency, i.e., at "resonance". This process is known as RF excitation and is the main mechanism of MRI signal generation.

After excitation with an RF pulse, the spin system experiences two concurrent processes known as relaxation. The longitudinal magnetisation along the magnetic field, M_z, gradually recovers. This process is called T1 relaxation, and its inverse is the relaxation rate, 1/T1 = R1. Similarly, the transverse magnetisation orthogonal to the main magnetic field, M_{xy}, decays due to loss of spin coherence. This gives rise to T2 relaxation and its relaxation rate, 1/T2 = R2. One might expect the T1 and T2 relaxation times to be identical, but microstructural effects and macroscale field inhomogeneities ensure that the T1 relaxation time is always greater than the T2 relaxation time in vivo. T1 and T2 relaxation times differ between different types of biological tissue, which is used to generate contrast between different tissue types.

Spatial information must also be encoded to form an MRI image. This is done by using magnetic gradients to introduce position-dependent variations in the Larmor frequency, enabling selective excitation of the tissue in all spatial directions. MRI pulse sequences are created by arranging RF excitation pulses and position-dependent gradients. This allows for the acquisition of measurements that are used to reconstruct the final MRI image. There is an abundance of different MRI pulse sequences available, and by varying their parameters, such as repetition time (TR) and echo time (TE), they can be used to create images with specific contrasts between tissue types. In the first two studies, we used T1-weighted (T1w; short TR and TE) and T2-weighted (T2w; long TR and TE) pulse sequences. T1w images provide excellent neuroanatomical contrast between grey and white matter and between brain and cerebrospinal fluid, whereas T2w images are frequently used to identify pathological processes that cause elevated concentrations of free water, which appear in the T2w images as hyperintensities.

In the third study, we used a diffusion-weighted pulse sequence. Such imaging techniques, collectively known as diffusion-weighted imaging (DWI), employ directional gradients to measure the MRI signal decay due to the diffusion of water molecules in biological tissue. In a homogeneous medium, this diffusion is random, represented by Brownian motion, and equal in all directions, i.e., isotropic. However, in biological tissue, water is contained within cells and extracellular compartments that restrict diffusion along specific directions, giving rise to anisotropic diffusion. In white matter, the primary sources of anisotropic diffusion are axons and the myelin enveloping them. Since the motion of protons induces phase shifts in the MRI signal, this gives rise to a signal that can be measured with DWI pulse sequences. This introduces diffusion sensitivity into the MRI signal, and the degree of diffusion weighting is determined by the b-factor, where higher b-values correspond to more diffusion weighting and greater sensitivity to slower diffusion processes. Importantly, DWI carries information about microstructural properties of the tissue and can be used to estimate the directionality of white matter fibre tracts and thus shed light on the structural connectivity of the brain.

1.2.2 The T1w/T2w-ratio

The T1w/T2w-ratio is formed by dividing T1w by T2w MRI images and has been proposed as a proxy measure of intracortical myelin.[98] This use of the T1w/T2w-ratio is based on several observations regarding the T1w and T2w signal in the cerebral cortex. First, the T1w signal can be used to identify and delineate regions according to the distribution of intracortical myelin both across the cerebral cortex and between cortical laminae.[99–102] Second, the T2w signal is roughly anticorrelated with myelin, so that the signal contributions from myelin are amplified by forming the ratio. Third, shared field inhomogeneities in the T1w and T2w images are correlated and therefore attenuated in the ratio.[98] Although the T1w/T2w-ratio is an indirect measure of intracortical myelin, it has the advantage of only requiring conventional MRI sequences, and as such it can be studied in a wide variety of settings. The correspondence between the T1w/T2w-ratio and myelin in white matter is, however, less clear.[103] For instance, a recent study found low correspondence between the T1w/T2w-ratio in the corpus callosum and histological measures of myelination.[104]

The use of the T1w/T2w-ratio as a measure of intracortical myelin can be placed within two broad categories. At the individual level, it can be used to delineate, or parcellate,[105] regions of the cerebral cortex according to their myeloarchitectonic properties. Notably, Glasser et al,[106] used the T1w/T2w-ratio together with functional activations recorded in a naturalistic story paradigm to parcellate the cerebral cortex into 180 regions. The second use of the T1w/T2w-ratio is for comparisons between individuals or assessing associations with other variables of interest, e.g., adolescent neurodevelopment,[107] cognitive measures,[108] or studying cortical demyelination in multiple sclerosis.[109] Importantly, parcellation relies on the within-subject spatial similarity between T1w/T2w-ratio values and the myeloarchitecture of the cerebral cortex, while group comparisons and associations with other variables of interest depend on the across-subject comparability of T1w/T2w-ratio values.

1.2.3 Diffusion tensor imaging

As discussed in section 1.2.1, DWI can be used to derive information about microscopic tissue properties and to map the fibre architecture of brain tissue. Diffusion models are used to mathematically represent the DWI signal,[110] and some can estimate a large number of parameters related to tissue microstructure. Still, these models typically require data acquired with DWI pulse sequences with long acquisition times that place high demands on scanner hardware. Diffusion tensor imaging (DTI) is a simpler model but it has the advantage of being easy to implement and is therefore commonly used in clinical research settings. DTI is primarily used to study white matter, as the cortical signal is weak due to low anisotropy. In particular, DTI metrics are thought to have low sensitivity to intracortical myelin.[111]

In DTI, the data is used to fit the diffusion tensor, represented by a 3-by-3 matrix for each voxel of the image. The matrices are characterised in terms of their eigenvalues and eigenvectors. Intuitively, the eigenvectors span an ellipsoid where the corresponding eigenvalues quantify the degree of diffusion along the principal axes. The four main DTI metrics are fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD).

DTI metrics reflect a combination of neurobiological processes, such as white matter fibre organisation and degree of myelination, and their sensitivities to different tissue properties differ. Briefly, FA alterations are frequently ascribed to fibre organisation, MD and RD to myelin-related alterations, and AD to axonal damage.[112,113] However, they are indirect measures of microstructure, and as such strong conclusions about their microstructural correlates should be avoided.[114] These and other challenges of in vivo imaging of myelin using MRI are discussed in the next section. See **Figure 3** for a depiction of the T1w/T2w-ratio and DTI metrics.

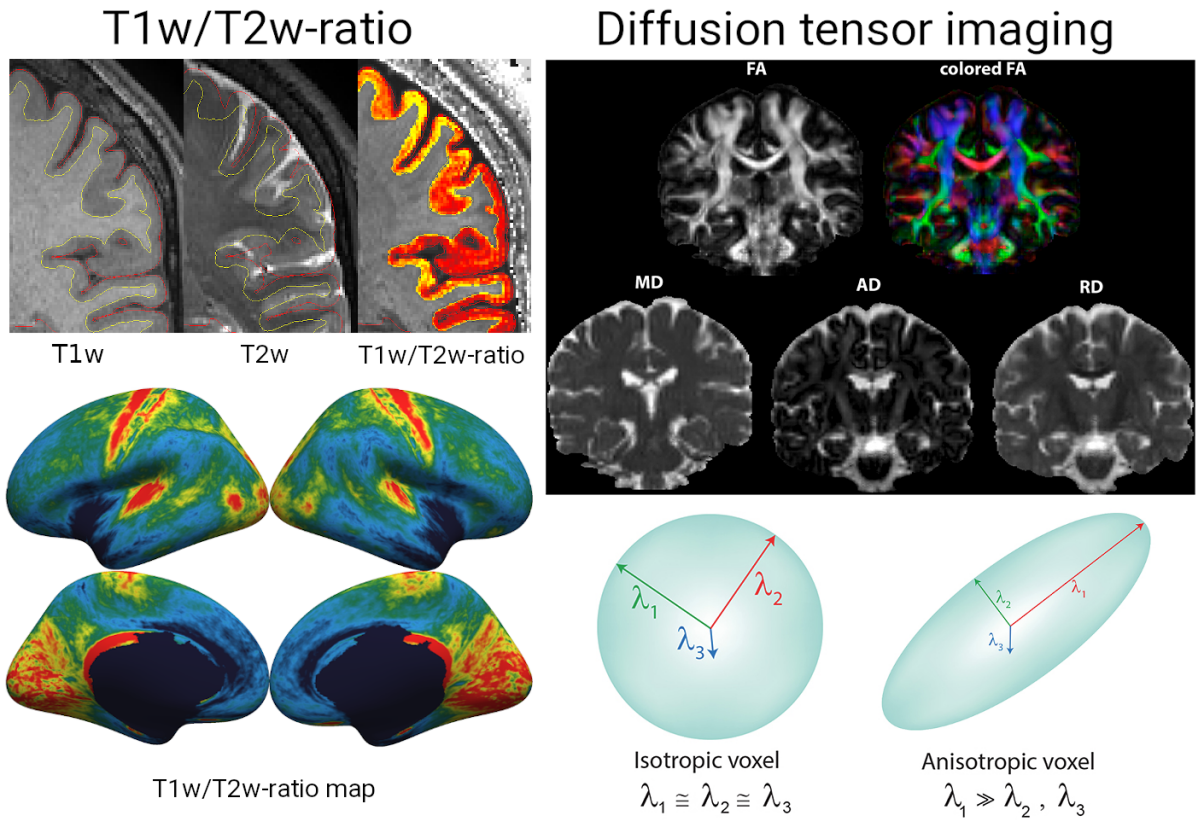


Figure 3. Left: The top images show how the T1w image is divided by the T2w image to form the T1w/T2w-ratio. The yellow curve represents the grey-white boundary and the red curve represents the boundary between cortex and cerebrospinal fluid. Yellow colours in the T1w/T2w-ratio represent higher T1w/T2w-ratio values. The bottom image depicts a mean T1w/T2w-ratio map on a standard cortical surface. Right: The top images show the different DTI metrics, FA, MD, RD, and AD. The bottom image illustrates how the eigenvalues of the diffusion tensor differ between an isotropic voxel with those of an anisotropic voxel. This figure is adapted from Moura et al.[115]

1.2.4 Challenges of in vivo myelin imaging

There are important methodological challenges in linking MRI measures of the brain to microstructure. In particular, myelin can impact the MRI signal in complex ways depending on tissue composition and structure.[116] Therefore, the use of appropriate MRI pulse sequences and tissue models is necessary to estimate myelin-related parameters. For example, DWI is more sensitive to myelin-related changes in white matter and less sensitive to intracortical myelin. Conversely, the T1w/T2w-ratio is thought to correspond closely to intracortical myelin but less so for white matter. It is important to note that other constituents of brain tissue besides myelin, such as paramagnetic iron content, also can influence the MRI signal. Furthermore, scanner model and pulse sequence parameters affect both image quality and MRI measures. Studying these effects is critical to ensure reproducible findings and facilitate their neurobiological interpretations.

To overcome these challenges, researchers have developed quantitative MRI (qMRI) techniques that combine multiple weighted MRI scans to estimate physical parameters of the measurement.[117] These techniques provide a more direct link to microstructure, including myelin-related parameters. For instance, the MP2RAGE pulse sequence, which estimates the T1 relaxation rate, R1, has been frequently used to study intracortical myelin, and the myelin water fraction (MWF), based on multi-echo T2 relaxation, is considered an in vivo measure of white matter myelination with high validity.[118] Similarly, the semi-quantitative magnetisation transfer (MT) pulse sequence can be used to compute the MT ratio (MTR), which corresponds closely with myelin.[119] Despite the advantages of qMRI and related techniques for inferring information about microstructural tissue properties, they often require advanced and time-consuming MRI pulse sequences, which has prevented widespread adoption. Therefore, this project focused on two imaging modalities related to intracortical and white matter myelination that are more commonly available in clinical studies.

The T1w/T2w-ratio is thought to reflect intracortical myelination, and high correlations with R1 have been reported in grey matter,[120] but the microstructural correlates of the T1w/T2w-ratio in white matter are less clear.[103,121] It is also unknown to what extent the T1w/T2w-ratio reflects intracortical myelin as opposed to cellular distributions that covary with myeloarchitecture. For instance, cellular bodies, such as those of oligodendrocytes and other glial cells, have high iron content,[122] which may affect the MRI signal. Notably, one study found stronger correlations between the T1w/T2w-ratio and expression of genes related to molecule size, oligodendrocytes, axon calibre, and pH than with myelin-related genes, although a relationship with genes related to myelination was also reported.[123] Finally, few studies to date have investigated the measurement properties of the T1w/T2w-ratio, such as test-retest reliability and the effects of scanner-related differences.

Similarly, there are unresolved questions regarding the relationship between DTI metrics and microstructure. For example, Björnholm et al[124] found that while R1, R2, and MWF corresponded closely with histological measures of white matter fibre diameter and density in the corpus callosum, FA and MD did not. Notably, the corpus callosum exhibits highly parallel fibre tract organisation compared to other fibre tracts. Effects due to more complex fibre tract organisation, e.g., fanning, bending, and crossing fibres, are known to affect DTI metrics in ways that render their interpretation in terms of microstructure more challenging.[125]

To address the intrinsic limitations of DTI, more sophisticated diffusion models have been developed. Such models include diffusion kurtosis imaging (DKI)[126] and neurite orientation dispersion and density imaging (NODDI).[127] These are based on multishell DWI where multiple b-values are used to acquire data with additional diffusion weightings. In addition, other imaging techniques exist that may be more specific to myelin. For instance, T2 relaxation time mapping is used to compute the MWF, which shows high correlations with histological myelin staining.[118,128] However, such sequences can be technically challenging to use and have long acquisition times. Their adoption, particularly in clinical studies, has therefore been slow.

In conclusion, the validity of MRI measures of myelin in grey and white matter remains to be determined, and more information is needed about their measurement properties. While proxy measures of myelin are frequently used in the study of schizophrenia spectrum and bipolar disorders, it is vital to be aware of methodological limitations when relating findings to microstructural properties and putative illness mechanisms.

1.3 Myelin imaging studies in schizophrenia spectrum and bipolar disorders

In vivo neuroimaging in schizophrenia spectrum and bipolar disorders have revealed a large number of findings that may reflect myelin-related alterations. Many studies have focused on white matter myelination, whereas fewer studies have explicitly focused on intracortical myelin. As discussed in section 1.1.6, histological studies in schizophrenia spectrum and bipolar disorders have identified ultrastructural and myelin-related abnormalities in patients. While such histological studies are immensely valuable, post-mortem analysis of brain tissue is time-consuming and necessarily limited to specific regions and small samples. It is also difficult in such studies to avoid selection bias and to investigate whether observed differences are causal or consequences of having lived with a severe mental disorder.

In vivo neuroimaging with MRI has the advantage in that it can include a large number of participants across the lifespan and can thus estimate longitudinal trajectories and better capture the clinical heterogeneity that characterises schizophrenia spectrum and bipolar disorders. In the next sections, we will present a selection of previous findings of myelin-related alterations in schizophrenia spectrum disorders, bipolar disorders, and associations between myelin-related measures and auditory hallucinations. This overview will necessarily be incomplete, given the multitude of studies on this topic.

1.3.1 Schizophrenia spectrum disorders

Numerous studies have reported alterations of DTI metrics in patients with schizophrenia spectrum disorders. The Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) consortium has pooled data across multiple sites worldwide to identify robust illness effects. In one such study, Kelly et al[129] compared 1963 patients with schizophrenia spectrum disorders and 2359 healthy controls and found widespread tract-wise differences in FA, MD, and RD. They found the largest effect sizes for FA within the entire white matter skeleton (d=0.42), the corona radiata (d=0.40), and the corpus callosum (d=0.39). Interestingly, these differences were more pronounced in peripheral regions, i.e., regions of the white matter skeleton outside of the Johns Hopkins University (JHU) white matter atlas,[130] rather than core regions that are contained within this atlas. Similar findings were reported in a study on early-onset psychosis conducted by our own group within the ENIGMA consortium,[131] and in a number of smaller studies,[132–134] with substantial heterogeneity in estimated effect sizes and the patterns of affected fibre tracts.[135]

Alterations in DTI metrics are often attributed to abnormal myelination. However, as discussed in section 1.2.4, it is difficult to directly link DTI metrics to specific features of white matter microstructure. Several studies have employed quantitative and semi-quantitative neuroimaging techniques in schizophrenia spectrum disorders. For instance, many studies report lower MTR in patients compared to healthy controls,[136,137] with some notable exceptions.[138] Lower MTR in the left cingulate has been associated with delusions of control in schizophrenia spectrum disorders.[139] In a study on MWF, Vanes et al[140,141] found reduced MWF in frontally projecting white matter fibre tracts in both chronic and recent-onset patients with schizophrenia spectrum disorders, as well as associations with IQ and verbal fluency in recent-onset patients. Using macromolecular proton fraction mapping, which is based on the MT effect,[142] Smirnova et al[143] found evidence of reduced myelination in both white and grey matter in patients with chronic schizophrenia spectrum disorders. Notably, the fraction was lower in patients characterised by negative symptoms and was negatively associated with the duration of illness.

Fewer studies have specifically investigated intracortical myelin in schizophrenia spectrum disorders. In a previous study, we examined the cortical grey-white matter contrast (GWC) in patients with schizophrenia spectrum and bipolar disorders.[144] The GWC is based on T1-weighted intensities in grey matter and adjacent superficial white matter and is thought to be inversely correlated with intracortical myelin content. We found higher GWC values in sensory and motor regions in 214 patients with schizophrenia spectrum disorders compared to 278 healthy controls, consistent with the interpretation of lower intracortical myelin in patients compared to healthy controls. More recently, Makowski et al[145] reported a trend-level increase in GWC values within a cortical component that included sensory and motor regions in patients with first-episode psychosis compared to healthy controls. Although this difference was not significant, the component contained regions overlapping with those where we found higher GWC values.

In an early study on the T1w/T2w-ratio, Iwatani et al[146] found lower mean, but not regional, T1w/T2w-ratio values in both grey and white matter in 29 patients with schizophrenia spectrum disorders and 33 healthy controls. Ganzetti et al[147] reported lower T1w/T2w-ratio values in the auditory cortex and in frontally projecting white matter fibre tracts in a comparison of 36 patients with schizophrenia spectrum disorder compared with 35 age-matched controls. More recently, Wei et al[148] investigated the T1w/T2w-ratio in patients with first-episode schizophrenia spectrum disorders with both direct group comparisons and structural covariance. They reported a more complex layer-dependent pattern of lower T1w/T2w-ratio values in the insula and posterior cingulate gyrus and higher values in the left inferior parietal lobe, supramarginal gyrus, and superior temporal gyrus. Interestingly, higher T1w/T2w-ratio values were associated with better cognitive function in younger participants but worse cognitive function in older participants. In another study using structural covariance, Wei et al[149] found both reduced and increased network-level T1w/T2w-ratio values and associations with duration of untreated psychosis.

It is important to consider the dynamic nature of myelination when interpreting the past literature on myelin-related differences in patients with schizophrenia spectrum disorders. The presence and severity of symptoms in these patients tend to fluctuate over time, and neurobiological correlates, insofar as they correlate with symptoms, would be expected to show similar fluctuations. Therefore, discrepancies in the literature may arise from the inclusion of patients at different stages of the illness. Importantly, treatment with antipsychotic medication may ameliorate myelin deficits, leading to null findings or paradoxical findings of higher myelin-related indices.[150,151] Hence, a comprehensive understanding of whether or not myelin is implicated in these disorders requires a nuanced and temporally sensitive approach.

1.3.2 Bipolar disorders

The DTI findings in bipolar disorders overlap substantially with those of schizophrenia spectrum disorders. Based on data from the ENIGMA consortium, Kochunov et al[152] showed high correlations in tract-wise effect sizes from patient-control comparisons in studies on schizophrenia spectrum disorders and bipolar disorders (r=0.72). Interestingly, they found high correlations between bipolar disorders and obsessive compulsive disorder (r=0.64), but not between bipolar disorders and major depressive disorder (r=0.28). Using DKI, Goghari et al[153] found lower mean and axial kurtosis in patients with bipolar disorders compared to healthy controls. While these findings may reflect abnormal myelination, it is difficult to link them directly to specific features of microstructure. It has been proposed that DWI findings in bipolar disorders reflect dysmyelination and that longitudinal changes in white matter microstructure correspond to the illness phases of bipolar disorders.[154]

Sehmbi et al[155–157] found associations between an indirect measure of intracortical myelin and verbal memory, as well as altered age trajectories of intracortical myelin in patients with bipolar I disorder. Structural covariance analyses showed associations between network measures and psychosocial functioning. Mahal et al[158] used magnetic resonance spectroscopy, an imaging technique for metabolic profiling, and found altered concentrations of choline metabolites in the dorsolateral prefrontal cortex in patients with bipolar I disorder.[158] They suggested that this reflects altered myelin turnover, as well as alterations in the neuronal membrane and overall cell density. In the previously mentioned study on the GWC from our own group,[144] we found higher GWC in 185 patients with bipolar disorders compared to 278 healthy controls. The implicated regions included the motor, auditory, and occipital cortices but were more limited than those of patients with schizophrenia spectrum disorders. Thus, there are findings congruent with alterations of both intracortical and white matter myelination in bipolar disorders. However, fewer studies have specifically investigated this putative pathophysiological mechanism in bipolar disorders compared to schizophrenia spectrum disorders.

1.3.3 Auditory hallucinations

Most previous studies on myelin-related MRI measures and auditory hallucinations have focused on schizophrenia spectrum disorders, whereas fewer studies have been conducted in bipolar disorders. DTI has been particularly frequently used in this context, and there are numerous reports of group differences in DTI metrics between patients with auditory hallucinations (AH+) and without auditory hallucinations (AH-), and between AH+, AH- and healthy controls.[159–161] It should also be pointed out that the assessment of auditory hallucination status varies between studies, with frequent use of both current and lifetime measures and a broad range of clinical instruments and criteria. We discuss these and related issues in more depth in section 5.2.5.

Previous studies have tended to focus on FA, and have primarily, or exclusively, investigated fibre tracts within the language and auditory processing circuitry (LAPC),[162] under the assumption that auditory hallucinations are specifically related to alterations of white matter microstructure within these fibre tracts. The most frequent findings involve lower FA in AH+ compared to healthy controls or AH-, but there is considerable heterogeneity in both the direction of findings and the precise fibre tracts that are implicated.[161] There has been a particular focus on the arcuate fasciculus (AF), given its important role in the LAPC.[163–166] Studies also report an effect of auditory hallucination status on DTI metrics in interhemispheric fibre tracts,[167–169] and it has been suggested that disrupted interhemispheric connectivity plays a role in auditory hallucinations.[170–172] A few DTI studies have also investigated a broader selection of DTI metrics, such as Sato et al,[173] who reported a positive association between hallucination severity and MD in individuals at ultra-high risk of developing schizophrenia and early-onset psychosis.

There have been few qMRI studies on auditory hallucinations focusing on myelin. However, an early qMRI study reported higher T1 relaxation times in the left temporal lobe in patients with epilepsy who had a clinical presentation similar to schizophrenia spectrum disorders.[174] Morphometric studies have shown an effect of auditory hallucination status on structural characteristics of the cerebral cortex. These findings include lower cortical thickness and grey matter density of auditory processing regions in AH+ compared to both AH- and healthy controls.[175] In bipolar disorders, the opposite finding of higher cortical thickness has been reported.[176] The microstructural correlates of group differences in cortical thickness are poorly understood, but it has been suggested that myelination may contribute to the measurement of cortical thickness by shifting the boundary between grey and white matter.[177,178] As such, morphometric findings may reflect myelin-related changes in superficial white or grey matter. In summary, previous studies on myelin-related MRI indices and auditory hallucinations have been mostly limited to DTI. These have tended to report lower FA and, to some extent, higher MD in AH+, whereas fewer studies have investigated the link between other myelin-related measures and auditory hallucinations.

1.3.4 Antipsychotic medication

As discussed in section 1.1.6, antipsychotic medication has been hypothesised to have promyelinating effects that may contribute to their clinical efficacy. To date, few comprehensive studies have investigated the effect of antipsychotic medication on DTI metrics.[179] Studies have mostly focused on FA, where both reduced[180] and increased[181–184] FA has been reported following antipsychotic treatment. Such studies are complicated by the fact that antipsychotic medication is the first line of treatment, can be a proxy for illness severity, and that cumulative information on antipsychotic medication use is usually not available.

In a series of studies, Bartzokis and colleagues investigated associations between antipsychotic medication, particularly atypical antipsychotic medications, and an indirect measure of intracortical myelin.[83,84,87,185] Taken together, their findings indicate a relationship between exposure to atypical antipsychotic medication and intracortical myelin. Initially, this relationship is positive but peaks after about one year and begins to decline.[87] They propose that initial promyelinating effects contribute to the reduction in positive symptoms, particularly within the first year of treatment, but are then attenuated, leading to an increased risk of remission. If confirmed, such a biological process would represent a promising target for the development of new medical treatments.[186,187]

It should be emphasised that these studies used an indirect measure of intracortical myelin. This measure is based on two MRI images with different sensitivity to myelinated cerebral cortex, i.e., a proton density and an inversion recovery image. White matter is segmented in both images, and the volumetric difference between the two segmentations yields a volumetric measure of intracortical myelin.[185] Given the indirect nature of this measure, future studies employing well-validated measures of intracortical myelin are needed to replicate their findings.

1.4 Summary

Schizophrenia and bipolar disorders are severe multifactorial mental disorders with heterogeneous clinical presentations and uncertain aetiology. Abnormal myelination has been advanced as a potential pathophysiological mechanism. Specifically, abnormal myelination may lead to a disruption of basic sensory processing that causes aberrant filtering of self-generated and externally generated perceptual information, giving rise to delusions of control and auditory hallucinations.

Although MRI is a remarkable tool for studying brain structure which is sensitive to myelin-related alterations, more knowledge is needed about the exact relationship between MRI imaging measures and cerebral microstructure, as well as other aspects of their measurement properties. This forms the background of the aims and objectives of the current doctoral project, which are summarised in the next chapter.

2. Aims and objectives

The aims of this doctoral project were to test two hypotheses in line with the dysmyelination hypothesis of psychosis. First, we hypothesised that intracortical myelin deficits are present in patients with schizophrenia spectrum and bipolar disorders. Second, we hypothesised that white matter myelination is affected in patients with schizophrenia spectrum disorders and auditory hallucinations.

To test these hypotheses, we first developed optimised pipelines for computing the T1w/T2w-ratio. Next, we used the optimised T1w/T2w-ratio to investigate group differences between patients with schizophrenia spectrum and bipolar disorders and healthy controls. Finally, we used DTI to assess how white matter microstructure is related to auditory hallucinations in schizophrenia spectrum disorders.

Objectives

1. Characterise the test-retest reliability of the T1w/T2w-ratio and its agreement with known myeloarchitecture.
2. Investigate whether optimised processing of the T1w/T2w-ratio can be used to improve its measurement properties.
3. Use the T1w/T2w-ratio to compare intracortical myelin in patients with schizophrenia spectrum and bipolar disorders with healthy controls.
4. Investigate if and how DTI metrics are associated with auditory hallucination status in patients with schizophrenia spectrum disorders.
5. Test whether putative alterations of DTI metrics in patients with auditory hallucinations are specific to fibre tracts within the LAPC.

Hypotheses

1. The T1w/T2w-ratio is influenced by the scanner model and pulse sequence and can be improved with the use of correction factors.
2. Patients, particularly those with schizophrenia spectrum disorders, have lower T1w/T2w-ratio values in primary motor and sensory cortices compared to healthy controls.
3. Patients with auditory hallucinations have lower FA and higher MD and RD compared to healthy controls in white matter fibre tracts within the LAPC.

3. Materials and methods

The doctoral project was conducted at the Norwegian Centre for Mental Disorders Research (NORMENT) and in collaboration with the Regional Core Facility in Translational MRI Neuroimaging at the Oslo University Hospital (OUS), Ullevål. In the third study, we included data from the Human Brain Informatics (HUBIN) project at Karolinska Institutet in Stockholm, Sweden.

The Thematically Organised Psychosis (TOP) study is a translational research project with the goal of studying the clinical and biological characteristics of psychotic disorders. It is organised within the NORMENT framework, which also coordinates several subprojects with a shared database and research protocols. Data collection for the TOP study began in 2002 and is still ongoing. The HUBIN project is a similar translational research initiative with a primary emphasis on schizophrenia spectrum disorders. Recruitment for the HUBIN project took place from 1998 to 2003, and follow-up examinations were conducted after 5 and 13 years.

NORMENT was established as a Norwegian Centre of Excellence in 2013 with a 10-year grant from the Research Council of Norway. It currently consists of 15 research groups and over 200 employees and affiliates. The research activities span a broad range of topics, including illness trajectories and outcomes, characterisation of symptomatology, analysis of genetic and epigenetic variation, research on pharmacology, as well as neuroimaging with electrophysiology and MRI.

3.1 Participants

In the first study, we included test-retest data from healthy participants scanned at our facility and two openly available datasets where test-retest data was available. In the second study, we included healthy controls and patients with schizophrenia spectrum and bipolar disorders from the TOP study. In the third study, we included patients with schizophrenia spectrum disorders and healthy controls from both the TOP study and from the HUBIN project.

3.1.1 Test-retest datasets

We included data from three test-retest datasets collected on four different MRI scanners. Two test-retest datasets, named Donders and HCP in the first study, were accessed through online repositories. The third dataset, named NOR-MR750 and NOR-Premier in the first study, was obtained in-house as part of a reliability study conducted in connection with an upgrade of the MRI scanner at the Regional Core Facility in Translational MRI Neuroimaging at OUS.

The Donders test-retest dataset included scans from 17 healthy participants (mean age=24.7 years) enrolled in a test-retest study at the Donders Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, Netherlands. The participants were scanned twice with repositioning between each MRI acquisition. Head fixation was randomised to cushions only or cushions and a chin rest. The participants were scanned on a 3T Siemens Magnetom Prisma scanner platform (Siemens, Erlangen, Germany) equipped with a 32-channel head coil. This dataset was previously used in a study by Shams, Norris and Marques,[120] where they compared the T1w/T2w-ratio with R1 mapping and assessed test-retest reliability.

The Human Connectome Project (HCP) test-retest dataset comprised acquisitions from participants (mean age=30.7 years) who took part in the HCP Young Adult (HCP-YA) study.[188] Participants were scanned twice on a customised 3T Siemens Connectom Skyra scanner platform (Siemens, Erlangen, Germany) equipped with a 32-channel receive head coil. Given the long scan-rescan interval of this dataset, we selected participants with a scan-rescan interval below 6 months, resulting in 34 participants with a mean scan-rescan interval of 3.85 months (range=[2-5]).

The in-house test-retest dataset consisted of scans from nine healthy participants (mean age= 35.8 years) recruited internally. The dataset was acquired as part of a scheduled upgrade of the MRI scanner platform from a 3T GE Discovery MR750 (GE Medical Systems, Milwaukee, USA) to a 3T GE SIGNA Premier. The aim of the study was to ensure data consistency across the old and new scanner platforms, and to optimise the new MRI protocol for T1w/T2w-ratio mapping. The study was planned and conducted by the doctoral candidate in collaboration with medical physicists Wibeke Nordhøy and Robin AB Bugge at the Department of Physics and Computational Radiology at OUS. We used a 2 × 2 × 2 design (scanner × scan session × protocol acquisition) with a total of 80 protocol acquisitions. We acquired complete data on 9 participants with similar age and sex distributions as participants in the TOP study.

The in-house test-retest dataset was designed to capture variation from multiple sources. First, the inclusion of two separate scan sessions accounted for the influence of day-to-day intra-subject variability. To minimise the effect of time-of-day variability, we additionally matched the time of day across sessions for each participant. Second, the study design allowed for both direct comparisons between MRI scanners and indirect comparisons of test-retest reliability between MRI scanners. In theory, it would be possible to also compare participants directly across scanner platform. Unfortunately, the upgrade was delayed and acquisitions on the upgraded scanner took place about 10 months after the first acquisitions. Each session included MRI scanning with a minimal protocol acquired twice with repositioning. See **Figure 4** for an overview of a study design.

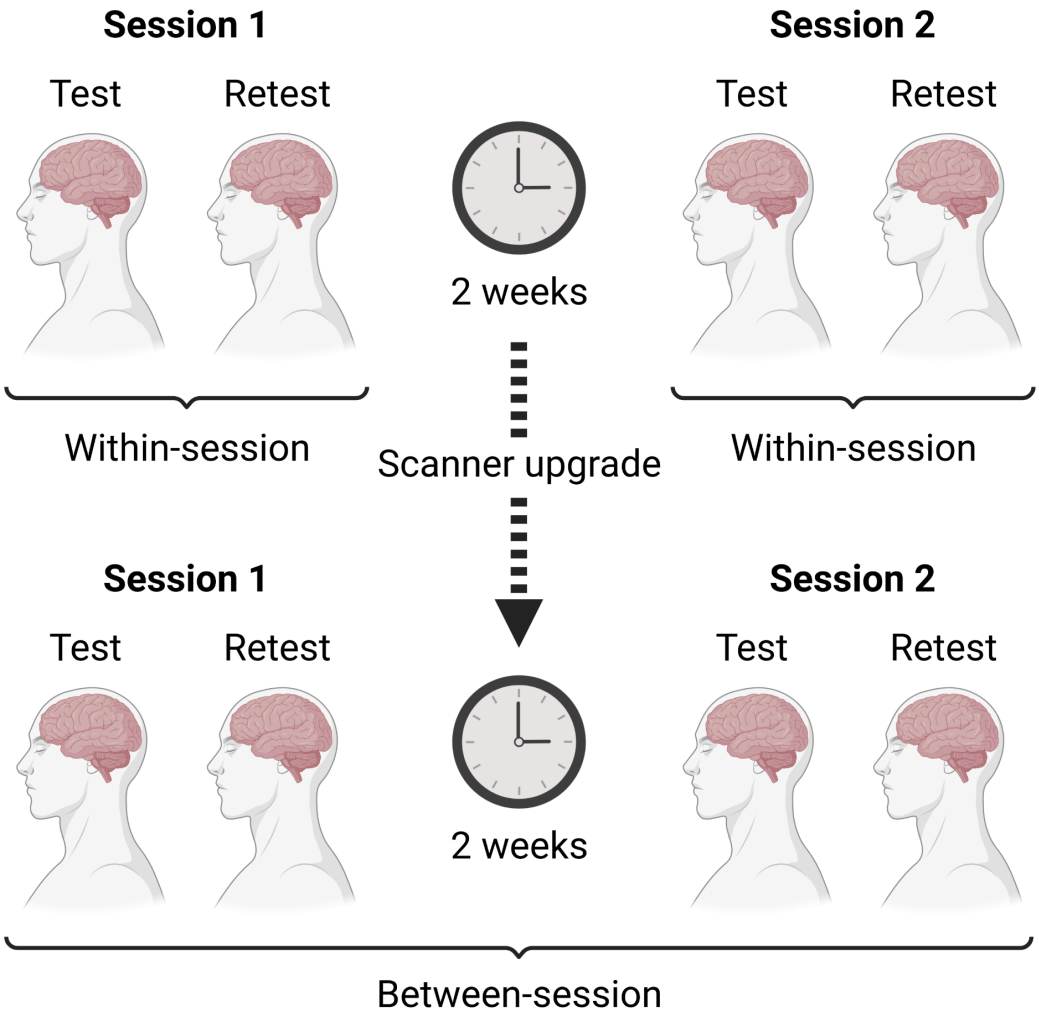


Figure 4. Overview of the in-house test-retest dataset. The top row depicts the two sessions with test-retest acquisitions on the Discovery MR750, and the bottom the acquisitions on the SIGNA Premier. Possible within- and between-session comparisons are illustrated.

3.1.2 TOP study

Patients with psychotic or affective symptoms were referred to participate in the TOP study by their clinicians at psychiatric inpatient ward units and outpatient clinics in the greater Oslo region. The patients underwent structured clinical assessment and were included if they were diagnosed with a DSM-IV disorder of schizophrenia spectrum (295.x), bipolar (296.x), or other psychotic disorders (297.x-298.x). Exclusion criteria included an age outside the range of 18-65 years, IQ less than 70, and a history of neurological illness or moderate to severe head injury. Patients with psychosis mainly due to an alcohol- or substance-induced disorder (291.x, 292.x) were excluded, but the presence of a comorbid alcohol or substance use disorder was not, in itself, an exclusion criteria for patients.

Healthy controls within the same age range (18-65 years) were randomly drawn from population registries in the same catchment area as the patients. Potential participants were contacted and asked to participate. Exclusion criteria for healthy controls were a mental disorder or a history of serious medical issues, a first-degree relative with a diagnosis of severe mental disorder, or meeting the diagnostic criteria for an alcohol or substance use disorder. Additionally, healthy controls were excluded if they reported cannabis use within the last three months.

In the second study, we included 310 participants from the TOP study, consisting of 64 patients with schizophrenia spectrum disorders, 91 patients with bipolar disorders, and 155 healthy controls. We only included participants scanned on the GE Discovery MR750 scanner platform since T2w images with sufficient quality and resolution were not included in previous MRI protocols.

In the third study, we included 228 participants from the TOP study, consisting of 114 patients with schizophrenia spectrum disorders scanned on both the GE Signa HDxt and GE Discovery MR750 scanner platforms. We selected patients with complete information on current auditory hallucinations.

3.1.3 HUBIN project

Patients with a preliminary diagnosis of a schizophrenia spectrum disorder were referred to the study by their treating psychiatrist within the catchment area of North-Western Stockholm County. For inclusion, they had to meet DSM-III-R or DSM-IV criteria for schizophrenia or schizoaffective disorder. Participants with an IQ less than 70, head injury resulting in loss of consciousness for more than five minutes, or a medical condition affecting the brain were excluded.

Healthy controls were recruited from population registries and among hospital staff and students at the psychiatric centres. Participants with a history of mental illness or first-degree relatives diagnosed with a psychotic disorder were excluded.

In the third study, we included 52 participants from the HUBIN project, consisting of 26 participants with schizophrenia spectrum disorders and 26 healthy controls. Due to the availability of DTI data, we selected participants scanned on the GE Discovery MR750 scanner platform. The patients had a mean duration of illness of 27.7 years and had been followed up for about 13 years from baseline. We further selected patients with complete information on current auditory hallucinations.

3.2 Ethical considerations

NORMENT collaborates with Bipolarforeningen Norge, and a user representative is engaged part-time in the TOP study. A user expert advisory group provides guidance on protocols, ethical questions, and public outreach. All studies were carried out in accordance with the Helsinki Declaration, and participants provided written informed consent. Participants may withdraw from the study at any time.

The TOP study is approved by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate, while the HUBIN project is approved by the Swedish Ethical Review Authority at Karolinska Institutet. External datasets had ethical approval from their respective local ethics committees. Data handling was compliant with Norwegian Data Protection Authority and GDPR regulations.

3.3 Clinical assessment

3.3.1 General clinical assessment

Diagnostic evaluations of patients in the TOP study were based on the Structured Clinical Interview for DSM-IV (SCID-IV),[189] while the SCID for DSM-III-R (SCID-III-R)[189] was used in the HUBIN project. Current antipsychotic medication use, including dosage and type, was determined through interviews and converted to chlorpromazine equivalent doses (CPZ; mg/day).[190] IQ was assessed with the Wechsler Abbreviated Scale of Intelligence (WASI-II)[191] in the TOP study and the Wechsler Adult Intelligence Scale (WAIS) in the HUBIN project. In the TOP study, psychosocial functioning was rated with the split version of the Global Assessment of Functioning Scale (GAF).[192] Age at onset was defined as the age at the first psychotic episode, verified using the SCID-IV or SCID-III-R. Duration of illness was calculated as years from age at onset to age at MRI scan. In the TOP study, alcohol and drug use was assessed with the Alcohol Use Disorder Identification Test (AUDIT)[193] and the Drug Use Disorders Identification Test (DUDIT).[194]

3.3.2 Psychotic symptom ratings

In the TOP study, lifetime symptoms were assessed with the SCID-IV and current symptoms, i.e., the week before clinical assessment, were assessed with the Positive and Negative Syndrome Scale, PANSS.[195] We converted PANSS scores to symptom factors from the consensus-based Wallwork five-factor model.[196] This model consists of five symptom dimensions: Positive (P1 + P3 + P5 + G9), negative (N1 + N2 + N3 + N4 + N6 + G7), disorganised (P2 + N5 + G11), excited (P4 + P7 + G8 + G14), and depressive (G2 + G3 + G6).

In the HUBIN project, lifetime symptoms were assessed with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN),[197] and current symptoms were evaluated with the Scales for the Assessment of Positive and Negative Symptoms, SAPS[198] and SANS.[199]

3.3.2 Auditory hallucination status

In the third study, we determined current auditory hallucination status using the PANSS-P3 item in TOP and the SAPS-H1 item in HUBIN. For TOP, individuals with a score of 3 or above on the PANSS-P3 were defined as having current auditory hallucinations (AH+), while those with a score below 3 were defined as not having current auditory hallucinations (AH-).

The lifetime auditory hallucinations (L-AH+) group was defined as individuals with a rating of threshold/present on the B16 item from the SCID, and the absence of lifetime auditory hallucinations (L-AH-) was defined as those with a rating of subthreshold/not present.

In HUBIN, individuals who scored 2 or above on the SAPS-H1 were defined as AH+, while AH- was defined as those scoring 1 or 0. The L-AH- group was defined as those who had a rating of absent/not clinically significant or mild/questionable on the 17.004 item of the SCAN. See Table 1 for an overview of assessment criteria for PANSS, SAPS, SCID, and SCAN.

	Instrument	Item	Item name	Description	Scoring
Current	**PANSS**	P3	Hallucinatory behaviour	Verbal report or behaviour indicating perceptions which are not generated by external stimuli. These may occur in the auditory, visual, olfactory or somatic realms.	1 = Absent 2 = Minimal 3 = Mild 4 = Moderate 5 = Moderate severe 6 = Severe 7 = Extreme
	SAPS	H1	Auditory Hallucinations	The patient reports voices, noises, or other sounds that no one else hears.	0 = None/Not at all 1 = Questionable 2 = Mild 3 = Moderate 4 = Marked 5 = Severe
Lifetime	**SCID**	B16	Auditory Hallucinations	Auditory hallucinations when fully awake, heard either inside or outside of the head.	1 = Absent or false 2 = Subthreshold 3 = Threshold or true
	SCAN	17.004	Frequency of verbal hallucinations	Presence of verbal auditory hallucinations.	0 = None 1 = Rarely 2 = Every week or so 3 = Most days 4 = Nearly all the time

Table 1. Criteria used to assess the presence of current (PANSS and SAPS) and lifetime auditory hallucinations (SCID and SCAN). Names, descriptions, and explanation of the scores are taken from the clinical instruments. Current symptoms included one week before clinical assessment.

3.4 MRI data processing

In this doctoral project, we employed two primary MRI brain phenotypes. The first two studies employed the T1w/T2w-ratio, where the first study was a comprehensive evaluation of techniques for computing this measure. In the third study, we used DTI metrics in white matter with particular focus on MD and RD, since these have been suggested as more sensitive to myelin-related alterations than FA and AD. We also included intracranial volume (ICV) as a potential confounding variable.

3.4.1 T1w/T2w-ratio processing

T1w images were first processed with the open source and freely available FreeSurfer software package,[200] where we used versions 6.0.0 and 7.1.0 for the first study, and version 6.0.0 for the second study. In the first study, we also used T2w images as inputs in the *recon-all* processing pipeline to refine the surface reconstruction. Briefly, FreeSurfer segments grey and white matter compartments and reconstructs surface meshes representing the boundaries between grey and white matter (grey-white surface) and between grey matter and cerebrospinal fluid/non-brain (pial surface).[201,202]

In the first study, 170 individual acquisitions were processed, comprised of nine participants in the in-house test-retest dataset, each scanned four times on each MRI scanner, 15 participants in the Donders test-retest dataset (after three participants had been excluded due to image quality), and 34 participants in the HCP test-retest dataset. In the second study, 220 individual acquisitions were processed.

Reconstructed surfaces were inspected by research assistants, who performed manual editing if reconstruction errors were identified, following standard FreeSurfer procedures.[203] T1w image quality was assessed and images were excluded if large artefacts were seen, e.g., due to motion. In the second study, we visually inspected each T1w/T2w-ratio map, if the maps deviated substantially from known myeloarchitecture, the T1w and T2w were inspected, and the participant was excluded if artefacts or low image quality were observed for either modality.

In the first study, we implemented and evaluated 33 pipelines for computing the T1w/T2w-ratio. These included nine pipelines with corrections for partial volume effects, surface outliers and intensity nonuniformities. We further evaluated eight different intensity normalisation methods implemented in the Intensity Normalization toolbox,[204,205] as well as the standard T1w/T2w-ratio pipeline from the HCP Minimal Processing Pipelines (HCP-MPP).[206] Briefly, we included and evaluated partial volume correction as described by Shafee et al,[207] surface-based outlier correction as described by Glasser et al,[98] and intensity nonuniformity correction using the classic N3[208] and the more recent N4ITK[209] algorithms.

For all processing methods, except HCP-MPP, we sampled T1w/T2w-ratio intensities in grey matter along equivolumetric surfaces at distances of 10% to 80% of cortical thickness from the grey-white surface.[210] We omitted distances of 0-10% and 80-100% to minimise partial volume and vascularisation effects, where intensities from superficial white matter, cerebrospinal fluid, or blood vessels may affect sampled T1w/T2w-ratio values within the cerebral cortex.[98,207]

In the second study, we computed the T1w/T2w-ratio using a pipeline that included corrections for partial volume effects, surface outliers, and intensity nonuniformities with the N4ITK algorithm, as well as intensity normalisation with WhiteStripe.[211] This pipeline produced T1w/T2w-ratio maps that corresponded well with known myeloarchitecture and showed good test-retest reliability in the first study. Importantly, WhiteStripe appeared to preserve interindividual variation, which is desirable in the context of between-subject comparisons.

We decided not to use *Z*-score intensity normalisation, which also showed good test-retest reliability. This intensity normalisation method is based on whole-brain intensities, whereas WhiteStripe only uses intensities within white matter.[212] Accordingly, we reasoned that WhiteStripe makes fewer assumptions about T1w and T2w values in the cerebral cortex, which was the region of interest (ROI). We discuss the benefits and drawbacks of intensity normalisation in section 5.2.3.

3.4.2 Diffusion tensor imaging and tractography

DTI data was processed using an optimised pipeline[213] and the FMRIB Software Library (FSL; version 6.0.3).[214] This pipeline includes corrections for noise,[215] Gibbs ringing,[216] and echo-planar imaging artefacts due to motion, eddy currents, and susceptibility distortions.[217,218] In a previous study, this pipeline showed substantially higher temporal signal-to-noise ratio than the original UK Biobank pipeline and was more sensitive to age-related associations.[213] To estimate the DTI metrics, FA, MD, RD, and AD, we used *dtifit* in FSL with the linear weighted least squares algorithm.

To reconstruct white matter fibre tracts, we performed diffusion tractography with the XTRACT[219] toolbox in FSL. This toolbox uses the *probtrackx2* command in FSL to produce streamlines that propagate probabilistically from a seed using information about fibre orientation. For each step in the propagation, termination criteria are checked, such as a streamline reaching its target or entering a region marked for exclusion. Similarly, waypoints can be specified that streamlines are required to pass through to be considered valid. This approach to isolating anatomically plausible fibre tracts using neuroanatomical priors in the form of seeds, exclusion, and waypoint masks is sometimes referred to as virtual dissection.[220]

XTRACT includes tractography protocols for 42 white matter fibre tracts, of which 19 are bilateral and four are commissural. The protocols include seeds, waypoints, exclusion masks, and target masks, defined in the MNI152 standard space that are mapped to subject space where the tractography takes place. Two seeding strategies are used depending on the fibre tract, either a single-ROI seed or a reverse-seeding approach, where a pair of seed-target masks are reversed and the final fibre tract is based on streamlines in both directions.[219]

Previous studies have highlighted low reproducibility as a methodological challenge with using diffusion tractography to reconstruct white matter fibre tracts.[221,222] The use of standardised tractography protocols has been proposed as one way to address this issue.[219] Furthermore, tractography performed at the subject-level may be more robust to individual neuroanatomical variability than registration-based approaches. We discuss these and related issues in section 5.2.6.

3.4.3 Intracranial volume

Since ICV has been shown to influence DTI metrics,[223] we included ICV as a possible confounding variable in the third study. To estimate ICV, we used the segmentation-based total intracranial volume (sbTIV) from the Sequence Adaptive Multimodal Segmentation (SAMSEG),[224] a toolbox included in FreeSurfer. It can be used with both T1w and T2w images as input, where we used T1w images only since T2w images were not available for some of the included datasets. Notably, SAMSEG has been designed to be robust with respect to differences in image quality and MRI pulse sequences. In a previous study, we compared five ICV estimation methods and their cross-sectional and longitudinal associations with age. In this comparison, SAMSEG showed less evidence of confounding by sex, height, weight, age, and total head size than the other ICV estimation methods.[225]

3.5 Statistical analyses

3.5.1 Reliability and agreement with myeloarchitecture of the T1w/T2w-ratio

Since MRI scanner and choice of pulse sequence may influence the T1w/T2w-ratio, we analysed each test-retest dataset separately. We first created mean T1w/T2w-ratio maps for each processing pipeline. These were compared qualitatively with an intracortical myelin map based on histological staining studies that was considered the gold standard. This map is based on studies from the Vogt-Vogt school and aligned to a standard brain by Nieuwenhuys and Broere. To quantify agreement with myeloarchitecture, we used two T1w/T2w-ratio reference maps, denoted YA-BC and YA-B1+, that correspond closely with histologically determined myeloarchitecture.[106,226]

To compare T1w/T2w-ratio maps with the reference maps, we first extracted mean T1w/T2w-ratio values in 176 ROIs from the HCP Multi-Modal Parcellation (HCP-MMP). These were used to compute Spearman rank correlations between each mean T1w/T2w-ratio map with the YA-BC and YA-B1+ reference maps. The reference maps are based on openly available datasets from the HCP-YA study,[188] where YA-BC was processed with the HCP-MPP pipeline and YA-B1+ was additionally corrected for B1 transmit (B1+) field inhomogeneities.[226] Note that we excluded four ROIs from the HCP-MMP due to proximity to the medial wall. We also briefly acknowledge that this was an indirect test of the agreement with known myeloarchitecture. This limitation is discussed in more depth in section 5.2.4.

We measured test-retest reliability by calculating percentage coefficients of variation (CVs) for cortical surface vertices, mean T1w/T2w-ratio values in the 176 HCP-MMP ROIs, and by combining vertex-wise CVs across the cortical surface. Furthermore, we extracted median T1w/T2w-ratio values across the cortical surface, which were used to compute intraclass correlation coefficients (ICCs) that quantify shifts of the subject-wise T1w/T2w-ratio distributions. We chose the median, since the mean is more sensitive to the presence of outliers, especially for the pipelines that did not include surface-based outlier correction. We further quantified the presence of lateral bias in T1w/T2w-ratio values by computing interhemispheric T1w/T2w-ratio differences for each ROI of the HCP-MMP with the percentage laterality index. Finally, we performed follow-up analyses to assess the effects of image resolution and data transformation from the standard surface used in the HCP-MPP to the standard surface used in FreeSurfer.

3.5.2 The T1w/T2w-ratio in schizophrenia spectrum and bipolar disorders

We first examined group differences between each patient group, i.e., schizophrenia spectrum and bipolar disorders, with healthy controls. To do this, we fitted vertex-wise regression models with diagnosis as a categorical predictor of interest and sex and age as covariates. T1w/T2w-ratio surface maps were concatenated and smoothed prior to the analyses. Separate contrasts were specified to compare each patient group with healthy controls. Cluster-wise correction for multiple testing was employed with a cluster-forming threshold set at 1×10^{-3} and cluster-wise probability of 5%, adjusted for comparisons across two cerebral hemispheres. Furthermore, we examined group differences in T1w/T2w-ratio age slopes. To do this, we specified diagnosis × age interaction contrasts to compare each patient group with healthy controls. We specified separate contrasts for each patient group with healthy controls. In other respects, the models were the same as those above.

In clusters where significant group differences were identified in the vertex-wise analyses, we extracted mean T1w/T2w-ratio values and performed follow-up analyses to examine the associations with antipsychotic medication use and psychotic symptoms. These analyses were performed for patients only, using regression models adjusted for age and sex. First, we fitted separate regression models where we included antipsychotic medication status (yes/no) and antipsychotic medication CPZ-equivalent dose as predictors of interest. Second, we assessed medication status-by-age and CPZ-by-age interactions. Finally, we performed exploratory analyses on the main effects of total PANSS scores and the five symptom dimensions of the Wallwork five-factor model. The cluster-wise analyses were corrected for multiple testing using the Benjamini-Hochberg procedure, where an adjusted p-value < 0.05 was considered statistically significant.

3.5.3 Auditory hallucinations and DTI in schizophrenia spectrum disorders

We fitted tract-wise regression models for each of the 39 included white matter fibre tracts, where we specified current auditory hallucination status as the independent variable of interest and each tract-wise DTI metric as dependent variables. The models were fitted for AH+, AH-, and healthy controls together for comparison with healthy controls, and among patients only for the direct comparison of AH+ with AH-. Similar analyses were run for lifetime auditory hallucination status. All regression models were adjusted for age, age^2, and sex.

Dimensionality reduction was conducted by using principal component analysis (PCA). Suitability of the data for PCA was assessed using the Kaiser-Meyer-Olkin test and Bartlett's test of sphericity. We created unimodal general factors (g-factors) of DTI metric-specific variation by pooling variation across all 39 fibre tracts for each DTI metric and extracting the first principal component, which was taken as the g-factor for that metric. This resulted in unimodal g-factors g-FA, g-MD, g-RD, and g-AD. Multimodal g-factors were created by conducting PCA on all tracts and DTI metrics simultaneously. Since prior studies indicated that two principal components are needed to capture multimodal variation, we extracted the first two principal components as our multimodal g-factors g-Dim1 and g-Dim2.

Finally, we tested group differences in uni- and multimodal g-factors between AH+, AH-, and healthy controls, as well as between AH+ and AH-. We further fitted regression models exploring associations between g-factors and age at onset, duration of illness, antipsychotic medication use (yes/no), CPZ, and Wallwork symptom factors. In these analyses, separate univariate regression models were fitted for each clinical measure as variables of interest, and with each uni- and multimodal g-factor as dependent variables.

In post-hoc analyses, we tested for confounding by ICV and body mass index (BMI) by fitting separate regression models adjusted for either ICV or BMI together with the current auditory hallucination status as the variable of interest. Age, age^2, and sex were included as covariates as above. We fitted the models in the whole sample to assess potential confounding in the comparison of g-factors for AH+ and AH- with those of healthy controls and among patients for the direct comparison between AH+ and AH-. The analyses were adjusted for multiple comparisons with the Benjamini-Hochberg procedure, and an adjusted p-value < 0.05 was considered statistically significant.

4. Results

4.1 Study 1: Reliability and agreement with myeloarchitecture of the T1w/T2w-ratio

In the first study, we included 33 pipelines for computing the T1w/T2w-ratio in 170 individual acquisitions of 58 participants. There were two main research questions. First, how closely does the T1w/T2w-ratio agree with known myeloarchitecture and can correction methods improve agreement? Second, what is the test-retest reliability of the T1w/T2w-ratio and can intensity normalisation improve reliability?

We found inconsistent agreement with myeloarchitecture across datasets and processing pipelines, which could in some cases be improved with intensity nonuniformity correction. Strikingly, the NOR-MR750 dataset had a Spearman rank correlation of only -0.05 with the YA-B1+ reference map for the raw T1w/T2w-ratio pipeline. However, this improved to 0.66 after N4ITK intensity nonuniformity correction and to 0.77 with the HCP-BC pipeline. Other datasets showed considerably higher correlations with this reference for the raw pipeline ranging from 0.72 for the HCP test-retest dataset to 0.96 for the Donders test-retest dataset.

The test-retest reliability was overall poor for the raw T1w/T2w-ratio, with CVs ranging from 12.3-8.6% and ICCs ranging from 0.13-0.90. Some intensity normalisation methods improved the reliability, whereas others paradoxically worsened it. We found large improvements for intensity normalisation with Z-score and WhiteStripe in combination with N4ITK intensity nonuniformity correction. Finally, we observed marginal effects of corrections for surface outliers and partial volume effects on both agreement with myeloarchitecture and test-retest reliability.

4.2 Study 2: The T1w/T2w-ratio in schizophrenia spectrum and bipolar disorders

In the second study, we included 64 patients with schizophrenia spectrum disorders, 91 patients with bipolar disorders, and 155 matched healthy controls. We used an optimised pipeline for computing T1w/T2w-ratio maps from the first study, which showed good test-retest reliability whilst preserving interindividual variation. We aimed to answer three research questions. First, does the T1w/T2w-ratio in patients differ compared to healthy controls? Second, do T1w/T2w-ratio age trajectories in patients differ compared to healthy controls? Third, is antipsychotic medication or psychotic symptoms related to T1w/T2w-ratio values or age trajectories in patients?

We observed no significant vertex-wise differences in the T1w/T2w-ratio between patients with schizophrenia spectrum or bipolar disorders and healthy controls. However, we found significant age-by-diagnosis interactions, indicating more positive age slopes of T1w/T2w-ratio values in patients with schizophrenia spectrum disorders compared to healthy controls.

Significant regions included clusters in frontal and temporal regions, including bilateral regions of the superior frontal and insular cortices, as well as parietal and occipital regions. Age slopes in healthy controls were negative in medial frontal and temporal regions with significant positive age associations limited to the central sulcus. In contrast, patients with schizophrenia spectrum disorders showed significant positive age associations in additional regions, including frontal regions. See **Figure 6** for an overview of significant clusters and scatterplots of T1w/T2w-ratio values against age for a selection of clusters.

In the cluster analyses restricted to the patients only, we found significant interactions between age and antipsychotic medication use (yes/no), indicating a more positive age slope in patients currently treated with antipsychotic medication. Furthermore, we found similar interaction effects with CPZ, where patients with a higher CPZ had more positive age slopes. The CPZ-by-age interaction was significant in all significant clusters except two out of 22. We observed no associations with psychotic symptoms that survived correction for multiple testing.

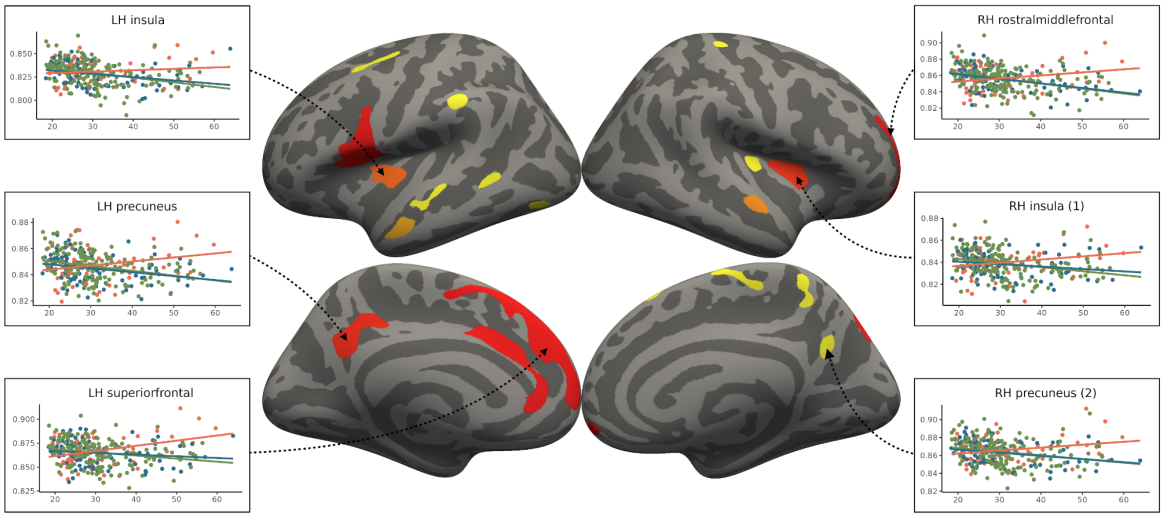


Figure 6. Clusters of significant age-by-diagnosis interactions in patients with schizophrenia spectrum disorders compared to healthy controls. Scatter plots show associations between age (x-axes) and T1w/T2w-ratio values (y-axes) for a selection of significant clusters among patients with schizophrenia spectrum disorders (red line), bipolar disorders (blue line) and healthy controls (green line).

4.3 Study 3: Auditory hallucinations and DTI in schizophrenia spectrum disorders

In the third study, we investigated the effect of auditory hallucination status on DTI metrics in 140 patients with schizophrenia spectrum disorders and 140 matched healthy controls. We aimed to answer three main questions. First, are DTI metrics different in AH+ compared to AH- or healthy controls, and is this effect specific to lower FA and higher MD and RD in fibre tracts of the LAPC? Second, can dimensionality reduction enhance sensitivity to the effects of auditory hallucination status? Third, are there associations with other psychotic symptoms or clinical factors such as antipsychotic medication and duration of illness?

Surprisingly, we found widespread tract-wise differences in MD and RD in the comparison of AH- with healthy controls, whereas differences between AH+ and healthy controls were limited to MD in the right middle longitudinal fasciculus (MDLF) and both MD and RD in the right optic radiation (OR). We did not observe any differences in the direct comparison of AH+ with AH-. For lifetime auditory hallucinations, we found tract-wise differences between both L-AH+ and L-AH- relative to healthy controls, where DTI alterations in L-AH- were mostly right lateralised and those of L-AH+ were left lateralised and bilateral. Only MD and RD differed significantly for these group comparisons. Importantly, the effects of auditory hallucination status were not restricted to fibre tracts within the LAPC.

In the dimensionality reduction analyses using PCA, we found that only AH- differed with respect to healthy controls. Significant differences were observed for all g-factors except g-FA. Specifically, g-AD was higher in AH- compared to healthy controls, despite no significant tract-wise findings for this DTI metric. There were no significant associations between the g-factors and age at onset, duration of illness, antipsychotic medication use or dose, or psychotic symptoms. ICV was significantly associated with g-FA, g-RD, g-AD, and g-Dim2, but group differences remained after adjusting for ICV or BMI. No significant differences were observed between AH+ and AH- for any of the g-factors. See **Figure 7** for an overview of the 39 white matter fibre tracts included in the analyses.

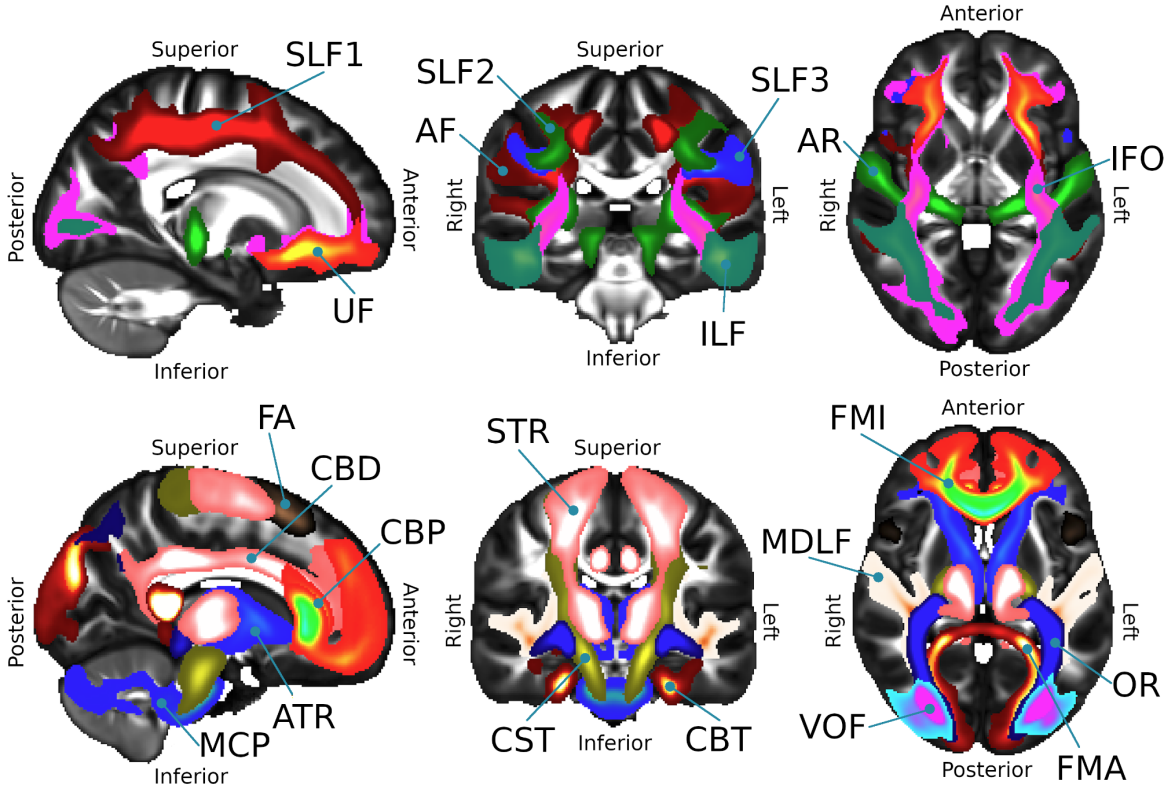


Figure 7. Overview of white matter fibre tracts included in the analyses in the third study. The top row shows fibre tracts within the language and auditory processing circuitry (LAPC) and the bottom row shows additional fibre tracts.

5. Discussion

In this doctoral project, we focused on two MRI measures related to myelin in grey and white matter respectively. We investigated methodological issues related to use of the T1w/T2w-ratio as a measure of intracortical myelin and aimed to test the dysmyelination hypothesis. In the first study, we performed a comprehensive evaluation of the test-retest reliability of the T1w/T2w-ratio and its agreement with known myeloarchitecture. Results from this study were then used to investigate the T1w/T2w-ratio in patients with schizophrenia spectrum and bipolar disorders. In the third and final study, we examined the relationship between auditory hallucination status and DTI metrics across a wide selection of white matter fibre tracts.

5.1 Main findings

5.1.1 Measurement properties of the T1w/T2w-ratio

The observation of poor agreement between T1w/T2w-ratio maps and known myeloarchitecture for some pipelines and datasets highlights the need for careful quality assurance and calibration of the T1w/T2w-ratio. This issue can be especially problematic when the T1w/T2w-ratio is used for cortical parcellation, which relies on cortical T1w/T2w-ratio distributions to define myeloarchitectonic boundaries between regions. We found that post hoc intensity nonuniformity correction can, in some cases, drastically improve this agreement. However, agreement ranged widely between the datasets, which likely reflects different homogeneity profiles of the MRI scanners. In particular, intensity inhomogeneities induced by the B1+ field (i.e., the B1 transmit field) can have deleterious effects on intensity-based measures. Such effects are highly nonlinear, have a major impact on weighted MRI intensities, and can be difficult to rectify with post hoc correction methods.[209]

In the dataset with the worst initial agreement with myeloarchitecture, we observed a radical improvement after intensity nonuniformity correction with the N4ITK algorithm. On the other hand, N4ITK paradoxically worsened the agreement in datasets where the raw T1w/T2w-ratio corresponded well with histologically determined myeloarchitecture and reference maps. We ascribe this effect to intensity nonuniformity correction getting 'greedy' if only minor field inhomogeneities are present in the data, leading to the introduction, rather than reduction, of intensity nonuniformities. Notably, this issue was more pronounced with N4ITK, which uses a more aggressive adjustment procedure compared to N3. As such, the decision whether or not to use post hoc intensity nonuniformity correction and, if used, the optimal choice of method depends on the presence and magnitude of field inhomogeneities in the dataset, which should be carefully assessed.

Theoretically, the effects of the receive (B1-) field are shared between the T1w and T2w images so that they cancel in the T1w/T2w-ratio.[98] B1+ field effects, on the other hand, will, in general, differ between the T1w and T2w images. They are therefore not necessarily cancelled in the T1w/T2w-ratio.[226] It is therefore particularly important to quantify intensity inhomogeneities related to the B1+ field and take steps to mitigate their effects. The gold standard of B1+ field mapping involves dedicated pulse sequences, which can be acquired in 1-2 minutes. However, many MRI protocols, particularly in clinical settings, do not include such field maps and indirect estimation of the B1+ field is therefore sometimes required. There are several approaches to reducing B1+ field inhomogeneities, including improved coil design and techniques such as B1+ shimming.[227] Notably, Glasser et al proposed two methods for mitigating B1+ field inhomogeneity effects when computing the T1w/T2w-ratio.[226] While one of the methods requires dedicated B1+ field maps, their pseudo-transmit approach instead uses two sequences that are commonly acquired in clinical MRI protocols. This method can therefore be an excellent alternative for datasets where such field maps have not been acquired.

Furthermore, our findings indicated low test-retest reliability for the T1w/T2w-ratio, observed across all datasets. Some intensity normalisation methods substantially improved the test-retest reliability. Importantly, the interindividual variation could be preserved with some intensity normalisation methods, which is desirable for group comparisons or assessing associations with other variables of interest. Surprisingly, we found that some intensity normalisation methods worsened test-retest reliability. This demonstrates the need to empirically assess the consequences of intensity normalisation to ensure that spurious variation is not introduced. As we will discuss further in the context of the second study, it is important to consider the consequences for interpretation when using intensity normalised data.

The low test-retest reliability may reflect transmit and receive gain settings determined during MRI acquisition. Since weighted MRI pulse sequences are typically optimised for contrast between specific tissue types and intraindividual disease markers, the interindividual comparability of intensities is frequently not a priority. Despite this, it may be feasible to design weighted MRI pulse sequences that are more appropriate for group comparisons. Another approach to ensuring consistency across acquisitions and scanner types is data acquisition with qMRI, which is an emerging field of research.[116,117,228–230] Historically, qMRI has been time-consuming and technically demanding, but in recent years qMRI pulse sequences have been developed that can be acquired within clinically acceptable time frames, and qMRI has become more suitable for clinical MRI protocols.[231]

Taken together, our findings indicate that agreement with myeloarchitecture is strongly dependent on scanner hardware and acquisition parameters but can be improved with the use of the appropriate correction methods. Furthermore, the test-retest reliability is overall low but can be improved with the use of some, but not all, intensity normalisation methods. We conclude that researchers employing the T1w/T2w-ratio should carefully assess their processing approach and be conscious of the intrinsic limitations of the T1w/T2w-ratio as a measure of intracortical myelin as well as in terms of its other measurement properties.

5.1.2 The T1w/T2w-ratio in schizophrenia spectrum and bipolar disorders

We did not observe any regional T1w/T2w-ratio differences between patients and healthy controls. This was not in agreement with our hypothesis and the findings from the previous study on the GWC, where we found significant differences in sensory and motor regions between patients with schizophrenia spectrum and bipolar disorders relative to healthy controls.[144] While the T1w/T2w-ratio and the GWC are different measures, they show moderate to high correlations and are both considered to be proxy measures of intracortical myelin.[232] Still, there have been few comprehensive studies on the measurement properties of the GWC, e.g., its dependence on MRI pulse sequence and scanner, and we cannot rule out that issues related to the reproducibility of the GWC affected our previous results.

The absence of a main effect of diagnosis on the T1w/T2w-ratio was also in disagreement with a previous study by Ganzetti et al,[147] where they found significantly reduced T1w/T2w-ratio in several cortical ROIs, including the temporal and frontal lobes and the insula. In this study, they performed intensity normalisation using intensities from masks covering the eyes and the temporal muscle.[233] However, their T2w images had a low resolution (4 mm slice thickness) which increases the risk of partial volume effects and may have affected their results. Furthermore, they pooled data from three different sites and given the large effects of the MRI scanner observed in the first study in this doctoral project, which may also have contributed to their findings. Our results are partially consistent with those of Iwatani et al,[146] who also did not see any significant regional group differences in the cerebral cortex. However, they excluded a large portion of the sensorimotor cortex from their analysis, and their low-resolution T2w images and large smoothing kernel used for computing the T1w/T2w-ratio could also introduce partial volume effects.

In two more recent studies, Wei et al[148,149] investigated the T1w/T2w-ratio in treatment naïve patients with first-episode schizophrenia spectrum disorders. They tested group differences between patients and healthy controls, as well as associations with psychotic symptoms and psychosocial functioning. They found a pattern of both higher and lower T1w/T2w-ratio values in patients relative to healthy controls, with higher values in the left superior gyrus and lower values in the left

cingulate and the insula. They used a similar equivolumetric surface projection method as we did, which is thought to correspond better to the laminar structure of the cerebral cortex.[210] Interestingly, they found that T1w/T2w-ratio values differed between patients and healthy controls for the middle and superficial layers, whereas the deep layer of the cortex showed no significant group differences. In our study, 89% of the patients with schizophrenia spectrum disorders were treated with antipsychotics, and they had a mean duration of illness of 7.2 years. Furthermore, Wei et al normalised T1w/T2w-ratio maps by demeaning and standardising the variance. Together, these methodological and clinical differences may have contributed to the apparent discrepancy with our findings.

Although we observed no main effect of diagnosis, we found different regional age trajectories in patients with schizophrenia spectrum disorders compared to healthy controls. These findings indicated less negative and more positive T1w/T2w-ratio values with increasing age. Intriguingly, there were interactions with antipsychotic medication use and dose. Specifically, patients currently treated with antipsychotics had more positive age trajectories with a dose-response relationship. These findings may be consistent with the hypothesised promyelinating effects of antipsychotic medication that were discussed in section 1.1.6.

As mentioned in the introduction, Bartzokis and colleagues have reported associations between treatment with atypical antipsychotics and indirect measures of intracortical myelin.[87,185] Preclinical studies have reported upregulation of cholesterol biosynthesis following administration of psychotropic drugs, and antipsychotics have been linked to serum lipid concentrations that are related to clinical efficacy.[88,90] There is therefore prior evidence to support the notion of a promyelinating effect of antipsychotics.

Given the cross-sectional and naturalistic study design, it was not possible to identify causal mechanisms. To test the hypothesis that antipsychotic medication affects intracortical myelination, measures sensitive to intracortical myelin should be acquired longitudinally for patients with schizophrenia spectrum disorders who are treatment naïve at baseline. Ideally, qMRI pulse sequences would be used to acquire data that is comparable across individuals and acquisitions. Such quantitative measures would strengthen the interpretation of any observed associations as related to intracortical myelin.

If our findings of divergent T1w/T2w-ratio age trajectories truly reflect antipsychotic medication treatment, such analyses could also be performed retrospectively in longitudinal studies where T1w and T2w pulse sequences are available. Nevertheless, the identification of neurobiologically processes specific to intracortical myelin would remain difficult and it can be argued that putative associations between the T1w/T2w-ratio and antipsychotic medication exposure relate to other changes, such as alterations in oligodendrocytes or paramagnetic iron concentrations.

5.1.3 Auditory hallucinations and DTI in schizophrenia spectrum disorders

In contrast to previous studies, we found only limited differences between patients with current auditory hallucinations (AH+) and healthy controls. On the other hand, we found a widespread pattern of elevated MD and RD in patients without current auditory hallucinations (AH-) relative to healthy controls. Higher MD and RD may be consistent with abnormal myelination, specifically lower white matter myelin content.

Unlike previous studies, we found no significant effect of auditory hallucination status on FA. Although many previous studies on auditory hallucinations in schizophrenia spectrum disorders have been limited by small sample sizes, there have also been well-powered studies reporting lower FA in AH+. We included 140 patients with schizophrenia spectrum disorders in our study, which is a similar or higher number than those of previous studies. We therefore consider it unlikely that we were underpowered to detect group differences in FA, and other explanations for the discrepancy appear more likely.

We found no significant associations between DTI metrics and duration of illness, age at onset, current antipsychotic medication use status or dose, or the five psychotic symptom dimensions. Nevertheless, we only had information on current antipsychotic use, and it is possible that cumulative exposure to antipsychotic medication would be associated with changes in white matter microstructure.

To our knowledge, there are few comprehensive studies on the effects of antipsychotic medication on white matter microstructure, and previous results have been inconsistent.[179–183] Based on the hypothesised promyelinating effect of antipsychotic medication, one could hypothesise that exposure to antipsychotics is associated with increased FA, as well as decreased MD and RD. On the other hand, studies on the relationship between antipsychotic exposure and intracortical myelin have reported a duration-dependent effect,[87] and a similar relationship may exist for white matter myelination. We encourage the use of comprehensive assessments of antipsychotic medication and preferable longitudinal study designs when assessing the effects of antipsychotic medication on white matter microstructure.

Another key finding was that the effects of auditory hallucination status on DTI metrics were not limited to the LAPC. Most previous studies have focused on these fibre tracts under the assumption that auditory hallucinations are specifically related to the LAPC. In the meta-analysis by Geoffroy et al from 2014,[234] they identified only five DTI studies that had compared patients with schizophrenia spectrum disorders and auditory verbal hallucinations with healthy controls. There was only sufficient data to perform a meta-analysis for the AF, a fibre tract that connects Wernicke and Broca's areas and plays an important role within the LAPC.[162,235]

Due to the sparsity of exploratory analyses on the relationship between auditory hallucinations and DTI metrics, it is difficult to form strong conclusions regarding the regional specificity of previous findings. Given the strong theoretical reasons for suspecting involvement of the LAPC,[160] there may be a bias towards such findings while incongruent findings are less often reported. To investigate this question, one could perform a large-scale analysis including multiple sites, e.g., via the ENIGMA consortium. This would allow for the estimation of publication bias in this literature,[236] as well as provide a robust test of the replicability of our findings.

Previous studies report a substantial degree of shared variation between DTI metrics across different white matter fibre tracts, which can be pooled to yield biologically interpretable general factors.[237–240] We found that dimensionality reduction with PCA enhanced sensitivity to differences in AD, which were not detected at the level of individual fibre tracts. Similarly, we found lower g-Dim2, a multimodal component with contributions mostly from AD (60.1%) and FA (33.9%). This may indicate that group differences in AD and FA are diffuse rather than tract-specific, which highlights the advantage of dimensionality reduction in such settings.

We found a trend towards lower FA in patients with schizophrenia spectrum disorders relative to healthy controls. Lower FA in this patient group is a commonly reported finding, notably in the large-scale ENIGMA analysis of 1963 patients with schizophrenia spectrum disorders,[129] yet there is considerable heterogeneity between studies, and methodological differences complicate comparison between studies.[241] Still, our results are mostly in line with previous results, and we consider it likely that the lack of significant differences in FA stem from methodological issues, such as the inclusion of a large number of fibre tracts.

For fibre tract reconstruction, we used a toolbox with robust protocols for reproducible diffusion tractography[219] and a unified processing framework based on an optimised pipeline.[213] Nevertheless, we note that given the large number of comparisons in the tract-wise comparisons, the precise fibre tracts that were implicated is likely to differ in a replication study. In particular, several fibre tracts were nominally significant before correction for multiple testing, and it is likely that different fibre tracts would be implicated depending on data quality and sample size. For that reason, we consider the main conclusion of the findings, as they relate to differences between AH- and healthy controls, to be one of widespread differences in MD and RD representing a general, rather than tract-specific, effect.

5.2 Methodological issues

5.2.1 The T1w/T2w-ratio as a measure of intracortical myelin

The T1w/T2w-ratio is a proxy measure of intracortical myelin and does not correspond directly to intracortical myelin content. While this is also true of qMRI measurements, such as T1 relaxation time mapping, these measures have a clearer biophysical interpretation. Parent et al[232] compared the T1w/T2w-ratio to the T1 relaxation rate, R1, as measured with the qMRI pulse sequence MP2RAGE. They found high correlations with the T1w/T2w-ratio sampled at a depth of 50% of cortical thickness. Similarly, Shams et al[120] found overall high correlations between R1 and the T1w/T2w-ratio. Importantly, the largest discrepancies between the two measures were observed in regions where the B1+ field deviated the most from its nominal value, consistent with our findings on the agreement with myeloarchitecture.

In many qMRI sequences, B1+ field inhomogeneities are adjusted for, either implicitly within the pulse sequence or retrospectively using B1+ field maps.[242] This is especially important when using MRI systems with high magnetic field strengths. However, such practices are not yet common practice for T1w/T2w-ratio mapping, and therefore issues pertaining to B1+ field inhomogeneities can be more problematic for the T1w/T2w-ratio. With a deeper appreciation of these issues and continued methodological development, there is considerable potential for improving the validity of the T1w/T2w-ratio as a measure of intracortical myelination.

It is also important to note that the relationship between the T1w/T2w-ratio and myelin is better established in grey compared to white matter. For example, Uddin et al[103,121] found low correlations between the T1w/T2w-ratio and the myelin water fraction (MWF) in white matter. That is not to say that the T1w/T2w-ratio cannot provide useful neurobiological information in white matter, but rather that the interpretation of T1w/T2w-ratio values as related to myelin is markedly more questionable than it is for the cerebral cortex.

In another comparison of MWF and the T1w/T2w-ratio, Arshad et al[243] reported similar findings of low correlations. In some regions, they found higher correlations between the T1w/T2w-ratio and a measure of T2 relaxation, the geometric mean of the relaxation decay signal. This led them to the conjecture that the T1w/T2w-ratio may be more sensitive to variation in axonal diameter and packing density rather than myelin content in white matter. However, the neurobiological interpretation of this T2 relaxation measure is also not well understood. Future studies should continue the work to clarify the link between the T1w/T2w-ratio and microstructural tissue properties, both in grey and white matter.

5.2.2 DTI metrics as a measure of white matter microstructure

As with the T1w/T2w-ratio and intracortical myelin, DTI metrics are proxy measures of white matter microstructure. There is an ongoing debate about their neurobiological correlates. For instance, the most studied DTI metric, FA, is often interpreted as 'white matter integrity'. However, as noted by Jones et al,[244] two voxels containing axons identical in every respect may have radically different FA due to the presence of crossing fibre tracts (low FA) rather than parallel fibre tracts (high FA). Importantly, such crossing fibres are present in about 60% of white matter voxels.[245]

Conversely, MD and RD are considered more sensitive to myelin-related alterations than FA, as reported by several studies on cuprizone-induced demyelination in mice.[112,246,247] At the same time, such animal models are likely to be associated with axonal damage and other tissue alterations, which makes it hard to form conclusions regarding the specificity to changes in myelination. Such studies also report time-dependent differences in the sensitivity of MD and RD to detect myelin-related changes, depending on the stage of de- and remyelination. It is also notable that one study found lower, rather than higher RD, following cuprizone-induced demyelination challenging the interpretation of higher RD as a marker of demyelination.[248]

As mentioned in the introduction, DTI is computed with a relatively simple diffusion model that has a nonspecific relationship with the tissue properties of white matter. More sophisticated diffusion models have been proposed that may provide better estimates of microstructural properties. Such models include DKI,[126] NODDI,[127] and white matter tract integrity (WMTI)[249,250] that are based on data from multishell DWI. Although these models provide a more accurate mathematical representation of the DWI signal than DTI, they are also subject to simplifying assumptions. Thus, their neurobiological correlates also remain to be determined, and their relationship with microstructure is an active field of research.[251]

5.2.3 Benefits and drawbacks of intensity normalisation

In the second study, we used intensity normalisation with WhiteStripe.[211] This method uses intensities in normal-appearing white matter (NAWM) to adjust T1w/T2w-ratio values in the cerebral cortex. It was chosen due to two desirable properties, namely it improved test-retest reliability whilst preserving interindividual variation, which is important for between-subject comparisons. Nonetheless, it is not without its downsides. Since it uses intensities in NAWM, it introduces a dependency between intensities in the cerebral cortex and intensities in white matter. It is possible that group differences in NAWM intensities can introduce or attenuate group differences in the cerebral cortex. We can therefore not rule out the possibility that the age-by-diagnosis interactions we found in the second study are not ultimately caused by age-by-diagnosis interactions within NAWM intensities.

Conversely, it is possible that intensity normalisation with WhiteStripe reduced sensitivity to group differences in the T1w/T2w-ratio. These possibilities are especially concerning, given previous reports of group differences in the T1w/T2w-ratio in white matter in patients with schizophrenia spectrum disorders relative to healthy controls.[146,147,252] Since the T1w and T2w signals are relative, it is challenging to address this shortcoming completely. As mentioned previously, qMRI pulse sequences can be one way to ensure that voxel values in the ROI can be directly compared between subjects without the need for intensity normalisation.

5.2.4 Comparing the T1w/T2w-ratio with known myeloarchitecture

In the first study, we compared T1w/T2w-ratio maps for each processing pipeline to two reference T1w/T2w-ratio maps, denoted YA-BC and YA-B1+. The rationale was that since the reference maps had been extensively validated against histological maps of myeloarchitecture, the agreement of our T1w/T2w-ratio maps with the reference maps would quantify agreement with myeloarchitecture. Although this was an indirect way to quantify the agreement, we considered it a valuable addition to the qualitative comparison with the myeloarchitectonic maps.

At the time of writing, there was, to our knowledge, no histology-based maps of myeloarchitecture that could be readily used for such a quantitative comparison. More recently, Foit et al[253] published a myeloarchitectonic atlas based on the histological Vogt-Vogt studies that is available for the standard fsaverage atlas. In future studies, we recommend that researchers employ this atlas, perhaps in addition to reference maps based on the T1w/T2w-ratio, to assess the correspondence of their T1w/T2w-ratio maps with known myeloarchitecture.

5.2.5 Assessing auditory hallucination status

Auditory hallucinations are considered core symptoms of schizophrenia spectrum disorders but are often insufficiently characterised in clinical studies. For example, the PANSS-P3 item includes hallucinations in visual, olfactory, and somatic modalities. Unlike the SAPS-H1 item, it is not specific to *auditory* hallucinations. As discussed in section 1.1.3, most individuals with psychotic disorders who report hallucinations in these modalities also report auditory hallucinations. Notably, Lim et al[28] assessed hallucinations in multiple modalities and found that 17.9% out of a total of 285 patients reported non-auditory hallucinations only within a month of clinical assessment. For lifetime hallucination measures, the proportion was 14.8% out of 603 patients. As such, the number of individuals reporting only non-auditory hallucinations is non-negligible, and the limited specificity of the PANSS-P3 item to auditory hallucinations must be acknowledged as a limitation.

In the HUBIN dataset, we assessed the co-occurrence of current auditory hallucinations and current hallucinations in other modalities. We found that 96.1% of the patients, i.e., all patients except one, with somatic, olfactory or visual hallucinations (i.e., a score above 1 on SAPS items H4, H5, and H6) also belonged to the AH+ group (i.e., a score above 2 on SAPS item H1). However, patients in the HUBIN dataset had a long duration of illness compared to those of the TOP study and it is unclear if this can be generalised to less chronic patient groups. It would be beneficial to include more comprehensive assessments of auditory hallucinations in future studies, especially since auditory hallucinations in psychotic disorders may be phenomenologically distinct from those in other groups.

5.2.6 Diffusion tractography and DTI analysis framework

The discrepant results that we saw in the third study may be explained in part by methodological differences. Notably, most previous studies on auditory hallucinations and DTI metrics have used tract-based spatial statistics (TBSS), whereas we used diffusion tractography implemented in the XTRACT toolbox. In TBSS,[254] FA images are mapped to a standard space using a nonlinear transformation. For each subject, the local maximum FA value along a skeleton in the standard space is then located and projected onto a common skeleton. The same is done for the other DTI metrics, MD, RD, and AD, where the projection is the same as for FA. The skeleton is then used either for voxel-wise comparisons between subjects or for extracting DTI metrics in predefined ROIs. As such, only a small region within each fibre tract is sampled in TBSS, in contrast to the approach we used, which was to extract the median of the DTI metrics within the entire fibre tract. Future studies are needed to assess the agreement between TBSS and XTRACT and to what extent these methods can lead to discrepant results.

6. Conclusions

In this doctoral project, we employed a two-step strategy for testing the dysmyelination hypothesis of psychosis. First, we performed a methodological appraisal of the T1w/T2w-ratio as a measure of intracortical myelin. Second, we performed clinical investigations of intracortical myelin using the T1w/T2w-ratio and DTI metrics for white matter myelination.

The methodological appraisal identified important limitations of the T1w/T2w-ratio as a measure of intracortical myelin. These limitations were partially mitigated with the appropriate processing methods. We found no evidence of intracortical myelin deficits in patients with schizophrenia spectrum and bipolar disorders. The observed divergent age trajectories may be consistent with promyelinating effects of antipsychotic medication. Furthermore, these effects may mask myelin deficits in these disorders. Current auditory hallucinations were not linked to white matter microstructure, but widespread differences were observed in patients without current auditory hallucinations compared to healthy controls. This suggests a more complex relationship between auditory hallucinations and white matter microstructure than previously thought. Future investigations would benefit from a comprehensive assessment of auditory hallucinations and a temporally sensitive approach.

Together the results from the three studies emphasised the need to be conscious of the merits and demerits of the measurements used for in vivo myelin imaging and how they affect the interpretations drawn from the data. Importantly, our results aligned with a growing body of evidence from multiple levels of observation indicating promyelinating effects of antipsychotic medication. Further studies employing rigorously tested and validated techniques for in vivo myelin imaging should be conducted to rigorously test the dysmyelination hypothesis. If a link between cerebral myelination and antipsychotic medication is established, this may represent a significant step forward in our understanding of how these medications achieve their therapeutic effect and could present promising targets for developing new medical treatments.

7. Acknowledgements

We express our gratitude to the study participants who participated in this project, the clinicians who were involved in recruitment and assessments at NORMENT and the HUBIN project, and to the radiographers and radiologists at Oslo University Hospital. A special thanks to the research assistants at the Imaging Psychosis lab, who have provided invaluable support throughout the years, Therese Stokkan, Ida Kippersund, Thomas Litleskare, Jannike Ausland, Viktoria Birkenæs, and Eirik Svela. We are also grateful for the innumerable people at NORMENT and Diakonhjemmet Hospital who make sure that the project progresses smoothly.

Data services were provided by the Services for Sensitive Data (TSD) platform, operated and developed by the TSD service group in the University of Oslo IT services department (USIT). We thank them for their efforts in maintaining this platform and for assisting whenever technical difficulties arise. A special thanks to the developers of the research software we used, both large projects such as FreeSurfer and FSL and small projects only used by a handful of researchers. Figures 1 and 4 were created by the doctoral candidate with BioRender.com.

The work was supported by The Research Council of Norway (grant numbers 223273, 274359), the K. G. Jebsen Foundation (grant number SKGJ-MED-008), the South-Eastern Norway Regional Health Authority (grant numbers 2017-097, 2019-104, 2020-020), the Swedish Research Council (K2012-61X-15078-09-3, K2015-62X-15077-12-3, 2017-00949), the regional agreement on medical training and clinical research between Stockholm County Council and the Karolinska Institutet, the Knut and Alice Wallenberg Foundation, and the HUBIN project.

8. Corrections

The word "not" was added to line 6 on page 16 so that the sentence now reads "... but **not** between bipolar disorders and …".

9. References

1. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9(2):137-150. doi:10.1016/S2215-0366(21)00395-3
2. Tandon R, Nasrallah HA, Keshavan MS. Schizophrenia, “just the facts” 4. Clinical features and conceptualization. *Schizophr Res*. 2009;110(1):1-23. doi:10.1016/j.schres.2009.03.005
3. Ali S, Santomauro D, Ferrari AJ, Charlson F. Excess mortality in severe mental disorders: A systematic review and meta-regression. *J Psychiatr Res*. 2022;149:97-105. doi:10.1016/j.jpsychires.2022.02.036
4. Laursen TM, Nordentoft M, Mortensen PB. Excess early mortality in schizophrenia. *Annu Rev Clin Psychol*. 2014;10:425-448. doi:10.1146/annurev-clinpsy-032813-153657
5. Collin G, Turk E, van den Heuvel MP. Connectomics in Schizophrenia: From Early Pioneers to Recent Brain Network Findings. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2016;1(3):199-208. doi:10.1016/j.bpsc.2016.01.002
6. Wernicke C. *Psychiatrie*. 1st ed. Leipzig: Barth; 1899.
7. Wernicke C. *Grundrisse Der Psychiatrie in Klinischen Vorlesungen*.; 1906.
8. Tamminga CA, Pearlson G, Keshavan M, Sweeney J, Clementz B, Thaker G. Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. *Schizophr Bull*. 2014;40 Suppl 2(Suppl 2):S131-137. doi:10.1093/schbul/sbt179
9. Pearlson GD. Etiologic, phenomenologic, and endophenotypic overlap of schizophrenia and bipolar disorder. *Annu Rev Clin Psychol*. 2015;11:251-281. doi:10.1146/annurev-clinpsy-032814-112915
10. Lieberman JA, First MB. Psychotic Disorders. *N Engl J Med*. 2018;379(3):270-280. doi:10.1056/NEJMra1801490
11. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752. doi:10.1038/nature08185
12. Ruderfer DM, Ripke S, McQuillin A, et al. Genomic dissection of bipolar disorder and schizophrenia including 28 subphenotypes. *Cell*. 2018;173(7):1705-1715.e16. doi:10.1016/j.cell.2018.05.046
13. Smeland OB, Bahrami S, Frei O, et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder and intelligence. *Mol Psychiatry*. January 2019:10.1038/s41380-018-0332-x. doi:10.1038/s41380-018-0332-x
14. van Bergen AH, Verkooijen S, Vreeker A, et al. The characteristics of psychotic features in bipolar disorder. *Psychol Med*. 2019;49(12):2036-2048. doi:10.1017/S0033291718002854
15. Aminoff SR, Onyeka IN, Ødegaard M, et al. Lifetime and point prevalence of psychotic symptoms in adults with bipolar disorders: a systematic review and meta-analysis. *Psychol Med*. 52(13):2413-2425. doi:10.1017/S003329172200201X
16. Chakrabarti S, Singh N. Psychotic symptoms in bipolar disorder and their impact on the illness: A systematic review. *World J Psychiatry*.

2022;12(9):1204-1232. doi:10.5498/wjp.v12.i9.1204
17. Perälä J, Suvisaari J, Saarni SI, et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Arch Gen Psychiatry*. 2007;64(1):19-28. doi:10.1001/archpsyc.64.1.19
18. Wunderink L, Sytema S, Nienhuis FJ, Wiersma D. Clinical recovery in first-episode psychosis. *Schizophr Bull*. 2009;35(2):362-369. doi:10.1093/schbul/sbn143
19. San L, Ciudad A, Alvarez E, Bobes J, Gilaberte I. Symptomatic remission and social/vocational functioning in outpatients with schizophrenia: prevalence and associations in a cross-sectional study. *Eur Psychiatry J Assoc Eur Psychiatr*. 2007;22(8):490-498. doi:10.1016/j.eurpsy.2007.06.005
20. Marwaha S, Johnson S, Bebbington P, et al. Rates and correlates of employment in people with schizophrenia in the UK, France and Germany. *Br J Psychiatry J Ment Sci*. 2007;191:30-37. doi:10.1192/bjp.bp.105.020982
21. Eack SM, Newhill CE. Psychiatric Symptoms and Quality of Life in Schizophrenia: A Meta-Analysis. *Schizophr Bull*. 2007;33(5):1225-1237. doi:10.1093/schbul/sbl071
22. Arnedo J, Svrakic DM, del Val C, et al. Uncovering the Hidden Risk Architecture of the Schizophrenias: Confirmation in Three Independent Genome-Wide Association Studies. *Am J Psychiatry*. 2015;172(2):139-153. doi:10.1176/appi.ajp.2014.14040435
23. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013;11:126. doi:10.1186/1741-7015-11-126
24. Merikangas KR, Jin R, He JP, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry*. 2011;68(3):241-251. doi:10.1001/archgenpsychiatry.2011.12
25. Keck PE, McElroy SL, Havens JR, et al. Psychosis in bipolar disorder: phenomenology and impact on morbidity and course of illness. *Compr Psychiatry*. 2003;44(4):263-269. doi:10.1016/S0010-440X(03)00089-0
26. Sartorius N, Jablensky A, Korten A, et al. Early manifestations and first-contact incidence of schizophrenia in different cultures. A preliminary report on the initial evaluation phase of the WHO Collaborative Study on determinants of outcome of severe mental disorders. *Psychol Med*. 1986;16(4):909-928. doi:10.1017/s0033291700011910
27. Goodwin FK, Jamison KR. *Manic-Depressive Illness. Bipolar Disorders and Recurrent Depression.* Oxford: Oxford University Press
28. Lim A, Hoek HW, Deen ML, et al. Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders. *Schizophr Res*. 2016;176(2):493-499. doi:10.1016/j.schres.2016.06.010
29. Nayani TH, David AS. The auditory hallucination: a phenomenological survey. *Psychol Med*. 1996;26(1):177-189. doi:10.1017/S003329170003381X
30. McCarthy-Jones S, Trauer T, Mackinnon A, Sims E, Thomas N, Copolov DL. A New Phenomenological Survey of Auditory Hallucinations: Evidence for Subtypes and Implications for Theory and Practice. *Schizophr Bull*. 2014;40(1):231-235. doi:10.1093/schbul/sbs156
31. Johns LC, Kompus K, Connell M, et al. Auditory verbal hallucinations in persons with and without a need for care. *Schizophr Bull*. 2014;40 Suppl 4(Suppl 4):S255-264. doi:10.1093/schbul/sbu005
32. Linszen MMJ, de Boer JN, Schutte MJL, et al. Occurrence and phenomenology of hallucinations in the general population: A large online survey.

Schizophr Heidelb Ger. 2022;8(1):41. doi:10.1038/s41537-022-00229-9
33. Larøi F. How do auditory verbal hallucinations in patients differ from those in non-patients? *Front Hum Neurosci*. 2012;6. https://www.frontiersin.org/articles/10.3389/fnhum.2012.00025. Accessed April 19, 2023.
34. Larøi F, Thomas N, Aleman A, et al. The ice in voices: Understanding negative content in auditory-verbal hallucinations. *Clin Psychol Rev*. 2019;67:1-10. doi:10.1016/j.cpr.2018.11.001
35. Feinberg I. Efference Copy and Corollary Discharge: Implications for Thinking and Its Disorders*. *Schizophr Bull*. 1978;4(4):636-640. doi:10.1093/schbul/4.4.636
36. Garrett M, Silva R. Auditory Hallucinations, Source Monitoring, and the Belief That “Voices” Are Real. *Schizophr Bull*. 2003;29(3):445-457. doi:10.1093/oxfordjournals.schbul.a007018
37. Moseley P, Fernyhough C, Ellison A. Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neurosci Biobehav Rev*. 2013;37(10):2794-2805. doi:10.1016/j.neubiorev.2013.10.001
38. Fernyhough C. Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas Psychol*. 2004;22(1):49-68. doi:10.1016/j.newideapsych.2004.09.001
39. Frith C. The neural basis of hallucinations and delusions. *C R Biol*. 2005;328(2):169-175. doi:10.1016/j.crvi.2004.10.012
40. Blakemore SJ, Rees G, Frith CD. How do we predict the consequences of our actions? A functional imaging study. *Neuropsychologia*. 1998;36(6):521-529. doi:10.1016/s0028-3932(97)00145-0
41. Wolpert DM. Computational approaches to motor control. *Trends Cogn Sci*. 1997;1(6):209-216. doi:10.1016/S1364-6613(97)01070-X
42. Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. *Science*. 1995;269(5232):1880-1882. doi:10.1126/science.7569931
43. Shergill SS, White TP, Joyce DW, Bays PM, Wolpert DM, Frith CD. Functional Magnetic Resonance Imaging of Impaired Sensory Prediction in Schizophrenia. *JAMA Psychiatry*. 2014;71(1):28-35. doi:10.1001/jamapsychiatry.2013.2974
44. Ford JM, Mathalon DH. Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? *Int J Psychophysiol*. 2005;58(2):179-189. doi:10.1016/j.ijpsycho.2005.01.014
45. Crapse TB, Sommer MA. Corollary discharge across the animal kingdom. *Nat Rev Neurosci*. 2008;9(8):587-600. doi:10.1038/nrn2457
46. Schneider DM, Nelson A, Mooney R. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*. 2014;513(7517):189-194. doi:10.1038/nature13724
47. Micheva KD, Wolman D, Mensh BD, et al. A large fraction of neocortical myelin ensheathes axons of local inhibitory neurons. Slutsky I, ed. *eLife*. 2016;5:e15784. doi:10.7554/eLife.15784
48. Javitt DC, Sweet RA. Auditory dysfunction in schizophrenia: integrating clinical and basic features. *Nat Rev Neurosci*. 2015;16(9):535-550. doi:10.1038/nrn4002
49. Schneider-Thoma J, Chalkou K, Dörries C, et al. Comparative efficacy and tolerability of 32 oral and long-acting injectable antipsychotics for the maintenance treatment of adults with schizophrenia: a systematic review and

network meta-analysis. *The Lancet*. 2022;399(10327):824-836. doi:10.1016/S0140-6736(21)01997-8
50. McCutcheon RA, Harrison PJ, Howes OD, McGuire PK, Taylor DM, Pillinger T. Data-driven Taxonomy for Antipsychotic Medication: A New Classification System. *Biol Psychiatry*. April 2023. doi:10.1016/j.biopsych.2023.04.004
51. McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, Dopamine and the Striatum: From Biology to Symptoms. *Trends Neurosci*. 2019;42(3):205-220. doi:10.1016/j.tins.2018.12.004
52. Emsley R, Chiliza B, Asmal L, Harvey BH. The nature of relapse in schizophrenia. *BMC Psychiatry*. 2013;13:50. doi:10.1186/1471-244X-13-50
53. Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr Bull*. 2009;35(3):549-562. doi:10.1093/schbul/sbp006
54. Friston K, Brown HR, Siemerkus J, Stephan KE. The dysconnection hypothesis (2016). *Schizophr Res*. 2016;176(2-3):83-94. doi:10.1016/j.schres.2016.07.014
55. McCutcheon RA, Krystal JH, Howes OD. Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry*. 2020;19(1):15-33. doi:10.1002/wps.20693
56. Murray RM, Bhavsar V, Tripoli G, Howes O. 30 Years on: How the Neurodevelopmental Hypothesis of Schizophrenia Morphed Into the Developmental Risk Factor Model of Psychosis. *Schizophr Bull*. 2017;43(6):1190-1196. doi:10.1093/schbul/sbx121
57. Lazari A, Salvan P, Verhagen L, et al. A macroscopic link between interhemispheric tract myelination and cortico-cortical interactions during action reprogramming. *Nat Commun*. 2022;13(1):4253. doi:10.1038/s41467-022-31687-5
58. Xin W, Chan JR. Myelin plasticity: sculpting circuits in learning and memory. *Nat Rev Neurosci*. 2020;21(12):682-694. doi:10.1038/s41583-020-00379-8
59. Watkins TA, Emery B, Mulinyawe S, Barres BA. Distinct stages of myelination regulated by gamma-secretase and astrocytes in a rapidly myelinating CNS coculture system. *Neuron*. 2008;60(4):555-569. doi:10.1016/j.neuron.2008.09.011
60. Funfschilling U, Supplie LM, Mahad D, et al. Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature*. 2012;485(7399):517-521. doi:10.1038/nature11007
61. Benamer N, Vidal M, Balia M, Angulo MC. Myelination of parvalbumin interneurons shapes the function of cortical sensory inhibitory circuits. *Nat Commun*. 2020;11:5151. doi:10.1038/s41467-020-18984-7
62. Grydeland H, Vértes PE, Váša F, et al. Waves of Maturation and Senescence in Micro-structural MRI Markers of Human Cortical Myelination over the Lifespan. *Cereb Cortex*. 2019;29(3):1369-1381. doi:10.1093/cercor/bhy330
63. Williamson JM, Lyons DA. Myelin Dynamics Throughout Life: An Ever-Changing Landscape? *Front Cell Neurosci*. 2018;12. https://www.frontiersin.org/articles/10.3389/fncel.2018.00424. Accessed April 20, 2023.
64. Chevalier N, Kurth S, Doucette MR, et al. Myelination Is Associated with Processing Speed in Early Childhood: Preliminary Insights. *PLoS ONE*. 2015;10(10):e0139897. doi:10.1371/journal.pone.0139897
65. Grydeland H, Westlye LT, Walhovd KB, Fjell AM. Intracortical Posterior

Cingulate Myelin Content Relates to Error Processing: Results from T1- and T2-Weighted MRI Myelin Mapping and Electrophysiology in Healthy Adults. *Cereb Cortex N Y N 1991*. 2016;26(6):2402-2410. doi:10.1093/cercor/bhv065
66. Corrigan NM, Yarnykh VL, Huber E, Zhao TC, Kuhl PK. Brain myelination at 7 months of age predicts later language development. *NeuroImage*. 2022;263:119641. doi:10.1016/j.neuroimage.2022.119641
67. Nieuwenhuys R, Broere CAJ. A detailed comparison of the cytoarchitectonic and myeloarchitectonic maps of the human neocortex produced by the Vogt–Vogt school. *Brain Struct Funct*. 2020;225(9):2717-2733. doi:10.1007/s00429-020-02150-2
68. Gray H. *Anatomy of the Human Body*. Philadelphia: Lea & Febiger; 1918.
69. Nieuwenhuys R, Broere CAJ. A map of the human neocortex showing the estimated overall myelin content of the individual architectonic areas based on the studies of Adolf Hopf. *Brain Struct Funct*. 2017;222(1):465-480. doi:10.1007/s00429-016-1228-7
70. Valdés-Tovar M, Rodríguez-Ramírez AM, Rodríguez-Cárdenas L, et al. Insights into myelin dysfunction in schizophrenia and bipolar disorder. *World J Psychiatry*. 2022;12(2):264-285. doi:10.5498/wjp.v12.i2.264
71. Stokowy T, Polushina T, Sønderby IE, et al. Genetic variation in 117 myelination-related genes in schizophrenia: Replication of association to lipid biosynthesis genes. *Sci Rep*. 2018;8(1):6915. doi:10.1038/s41598-018-25280-4
72. Hakak Y, Walker JR, Li C, et al. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc Natl Acad Sci U S A*. 2001;98(8):4746-4751. doi:10.1073/pnas.081071198
73. Katsel P, Davis KL, Haroutunian V. Variations in myelin and oligodendrocyte-related gene expression across multiple brain regions in schizophrenia: a gene ontology study. *Schizophr Res*. 2005;79(2-3):157-173. doi:10.1016/j.schres.2005.06.007
74. Sugai T, Kawamura M, Iritani S, et al. Prefrontal abnormality of schizophrenia revealed by DNA microarray: impact on glial and neurotrophic gene expression. *Ann N Y Acad Sci*. 2004;1025:84-91. doi:10.1196/annals.1316.011
75. Duncan LE, Holmans PA, Lee PH, et al. Pathway analyses implicate glial cells in schizophrenia. *PloS One*. 2014;9(2):e89441. doi:10.1371/journal.pone.0089441
76. Uranova NA, Vikhreva OV, Rachmanova VI, Orlovskaya DD. Ultrastructural alterations of myelinated fibers and oligodendrocytes in the prefrontal cortex in schizophrenia: a postmortem morphometric study. *Schizophr Res Treat*. 2011;2011:325789. doi:10.1155/2011/325789
77. Vostrikov VM, Uranova NA. Reduced density of oligodendrocytes and oligodendrocyte clusters in the caudate nucleus in major psychiatric illnesses. *Schizophr Res*. 2020;215:211-216. doi:10.1016/j.schres.2019.10.027
78. Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI. Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. *Schizophr Res*. 2004;67(2-3):269-275. doi:10.1016/S0920-9964(03)00181-6
79. Lake EMR, Steffler EA, Rowley CD, et al. Altered intracortical myelin staining in the dorsolateral prefrontal cortex in severe mental illness. *Eur Arch Psychiatry Clin Neurosci*. 2017;267(5):369-376. doi:10.1007/s00406-016-0730-5
80. Regenold WT, Phatak P, Marano CM, Gearhart L, Viens CH, Hisley KC. Myelin staining of deep white matter in the dorsolateral prefrontal cortex in

schizophrenia, bipolar disorder, and unipolar major depression. *Psychiatry Res*. 2007;151(3):179-188. doi:10.1016/j.psychres.2006.12.019
81. Insel TR. Rethinking schizophrenia. *Nature*. 2010;468(7321):187-193. doi:10.1038/nature09552
82. Whitford TJ, Ford JM, Mathalon DH, Kubicki M, Shenton ME. Schizophrenia, Myelination, and Delayed Corollary Discharges: A Hypothesis. *Schizophr Bull*. 2012;38(3):486-494. doi:10.1093/schbul/sbq105
83. Bartzokis G. Neuroglialpharmacology: Myelination As A Shared Mechanism of Action of Psychotropic Treatments. *Neuropharmacology*. 2012;62(7):2137-2153. doi:10.1016/j.neuropharm.2012.01.015
84. Bartzokis G, Lu PH, Stewart SB, et al. In vivo evidence of differential impact of typical and atypical antipsychotics on intracortical myelin in adults with schizophrenia. *Schizophr Res*. 2009;113(2):322-331. doi:10.1016/j.schres.2009.06.014
85. Bartzokis G. Schizophrenia: breakdown in the well-regulated lifelong process of brain development and maturation. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol*. 2002;27(4):672-683. doi:10.1016/S0893-133X(02)00364-0
86. Bartzokis G, Altshuler L. Reduced intracortical myelination in schizophrenia. *Am J Psychiatry*. 2005;162(6):1229-1230. doi:10.1176/appi.ajp.162.6.1229-a
87. Tishler TA, Bartzokis G, Lu PH, et al. Abnormal Trajectory of Intracortical Myelination in Schizophrenia Implicates White Matter in Disease Pathophysiology and the Therapeutic Mechanism of Action of Antipsychotics. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2018;3(5):454-462. doi:10.1016/j.bpsc.2017.03.007
88. Procyshyn RM, Wasan KM, Thornton AE, et al. Changes in serum lipids, independent of weight, are associated with changes in symptoms during long-term clozapine treatment. *J Psychiatry Neurosci JPN*. 2007;32(5):331-338.
89. Gjerde PB, Dieset I, Simonsen C, et al. Increase in serum HDL level is associated with less negative symptoms after one year of antipsychotic treatment in first-episode psychosis. *Schizophr Res*. 2018;197:253-260. doi:10.1016/j.schres.2017.10.042
90. Kim DD, Barr AM, Fredrikson DH, Honer WG, Procyshyn RM. Association between Serum Lipids and Antipsychotic Response in Schizophrenia. *Curr Neuropharmacol*. 2019;17(9):852-860. doi:10.2174/1570159X17666190228113348
91. Kroken RA, Løberg EM, Drønen T, et al. A Critical Review of Pro-Cognitive Drug Targets in Psychosis: Convergence on Myelination and Inflammation. *Front Psychiatry*. 2014;5:11. doi:10.3389/fpsyt.2014.00011
92. Ferno J, Skrede S, Vik-Mo AO, Jassim G, Le Hellard S, Steen VM. Lipogenic effects of psychotropic drugs: focus on the SREBP system. *Front Biosci Landmark Ed*. 2011;16(1):49-60. doi:10.2741/3675
93. Ersland KM, Skrede S, Stansberg C, Steen VM. Subchronic olanzapine exposure leads to increased expression of myelination-related genes in rat fronto-medial cortex. *Transl Psychiatry*. 2017;7:1262. doi:10.1038/s41398-017-0008-3
94. Mihai DP, Ungurianu A, Ciotu CI, et al. Effects of Venlafaxine, Risperidone and Febuxostat on Cuprizone-Induced Demyelination, Behavioral Deficits and Oxidative Stress. *Int J Mol Sci*. 2021;22(13):7183. doi:10.3390/ijms22137183
95. Zhou CH, Xue SS, Xue F, et al. The impact of quetiapine on the brain lipidome

in a cuprizone-induced mouse model of schizophrenia. *Biomed Pharmacother Biomedecine Pharmacother*. 2020;131:110707. doi:10.1016/j.biopha.2020.110707
96. Templeton N, Kivell B, McCaughey-Chapman A, Connor B, La Flamme AC. Clozapine administration enhanced functional recovery after cuprizone demyelination. *PloS One*. 2019;14(5):e0216113. doi:10.1371/journal.pone.0216113
97. Voineskos AN, Felsky D, Kovacevic N, et al. Oligodendrocyte genes, white matter tract integrity, and cognition in schizophrenia. *Cereb Cortex N Y N 1991*. 2013;23(9):2044-2057. doi:10.1093/cercor/bhs188
98. Glasser MF, Essen DCV. Mapping Human Cortical Areas In Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI. *J Neurosci*. 2011;31(32):11597-11616. doi:10.1523/JNEUROSCI.2180-11.2011
99. Clark VP, Courchesne E, Grafe M. In vivo Myeloarchitectonic Analysis of Human Striate and Extrastriate Cortex Using Magnetic Resonance Imaging. *Cereb Cortex*. 1992;2(5):417-424. doi:10.1093/cercor/2.5.417
100. Eickhoff S, Walters NB, Schleicher A, et al. High‐resolution MRI reflects myeloarchitecture and cytoarchitecture of human cerebral cortex. *Hum Brain Mapp*. 2004;24(3):206-215. doi:10.1002/hbm.20082
101. Walters NB, Egan GF, Kril JJ, et al. In vivo identification of human cortical areas using high-resolution MRI: An approach to cerebral structure–function correlation. *Proc Natl Acad Sci U S A*. 2003;100(5):2981-2986. doi:10.1073/pnas.0437896100
102. Koenig SH. Cholesterol of myelin is the determinant of gray-white contrast in MRI of brain. *Magn Reson Med*. 1991;20(2):285-291. doi:https://doi.org/10.1002/mrm.1910200210
103. Uddin MN, Figley TD, Marrie RA, Figley CR. Can T1w/T2w ratio be used as a myelin-specific measure in subcortical structures? Comparisons between FSE-based T1w/T2w ratios, GRASE-based T1w/T2w ratios and multi-echo GRASE-based myelin water fractions. *NMR Biomed*. 2018;31(3):e3868. doi:10.1002/nbm.3868
104. Sandrone S, Aiello M, Cavaliere C, et al. Mapping myelin in white matter with T1-weighted/T2-weighted maps: discrepancy with histology and other myelin MRI measures. *Brain Struct Funct*. 2023;228(2):525-535. doi:10.1007/s00429-022-02600-z
105. Eickhoff SB, Yeo BTT, Genon S. Imaging-based parcellations of the human brain. *Nat Rev Neurosci*. 2018;19(11):672-686. doi:10.1038/s41583-018-0071-7
106. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016;536(7615):171-178. doi:10.1038/nature18933
107. Norbom LB, Rokicki J, Alnæs D, et al. Maturation of cortical microstructure and cognitive development in childhood and adolescence: A T1w/T2w ratio MRI study. *Hum Brain Mapp*. 2020;41(16):4676-4690. doi:10.1002/hbm.25149
108. Grydeland H, Walhovd KB, Tamnes CK, Westlye LT, Fjell AM. Intracortical Myelin Links with Performance Variability across the Human Lifespan: Results from T1- and T2-Weighted MRI Myelin Mapping and Diffusion Tensor Imaging. *J Neurosci*. 2013;33(47):18618-18630. doi:10.1523/JNEUROSCI.2811-13.2013
109. Nakamura K, Chen JT, Ontaneda D, Fox RJ, Trapp BD. T1-/T2-weighted ratio differs in demyelinated cortex in multiple sclerosis. *Ann Neurol*. 2017;82(4):635-639. doi:10.1002/ana.25019

110. Jelescu IO, Palombo M, Bagnato F, Schilling KG. Challenges for biophysical modeling of microstructure. *J Neurosci Methods*. 2020;344:108861. doi:10.1016/j.jneumeth.2020.108861
111. Reveley C, Ye FQ, Mars RB, Matrov D, Chudasama Y, Leopold DA. Diffusion MRI anisotropy in the cerebral cortex is determined by unmyelinated tissue features. *Nat Commun*. 2022;13(1):6702. doi:10.1038/s41467-022-34328-z
112. Sun SW, Liang HF, Trinkaus K, Cross AH, Armstrong RC, Song SK. Noninvasive detection of cuprizone induced axonal damage and demyelination in the mouse corpus callosum. *Magn Reson Med*. 2006;55(2):302-308. doi:10.1002/mrm.20774
113. Peters JM, Struyven RR, Prohl AK, et al. White matter mean diffusivity correlates with myelination in tuberous sclerosis complex. *Ann Clin Transl Neurol*. 2019;6(7):1178-1190. doi:10.1002/acn3.793
114. Figley CR, Uddin MN, Wong K, Kornelsen J, Puig J, Figley TD. Potential Pitfalls of Using Fractional Anisotropy, Axial Diffusivity, and Radial Diffusivity as Biomarkers of Cerebral White Matter Microstructure. *Front Neurosci*. 2022;15. https://www.frontiersin.org/articles/10.3389/fnins.2021.799576. Accessed January 25, 2023.
115. Moura LM, Luccas R, Paiva JPQ de, et al. Diffusion Tensor Imaging Biomarkers to Predict Motor Outcomes in Stroke: A Narrative Review. *Front Neurol*. 2019;10. https://www.frontiersin.org/articles/10.3389/fneur.2019.00445. Accessed May 15, 2023.
116. Edwards LJ, Kirilina E, Mohammadi S, Weiskopf N. Microstructural imaging of human neocortex in vivo. *NeuroImage*. 2018;182:184-206. doi:10.1016/j.neuroimage.2018.02.055
117. Weiskopf N, Edwards LJ, Helms G, Mohammadi S, Kirilina E. Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nat Rev Phys*. 2021;3(8):570-588. doi:10.1038/s42254-021-00326-1
118. MacKay AL, Laule C. Magnetic Resonance of Myelin Water: An in vivo Marker for Myelin. *Brain Plast*. 2(1):71-91. doi:10.3233/BPL-160033
119. Stikov N, Perry LM, Mezer A, et al. Bound pool fractions complement diffusion measures to describe white matter micro and macrostructure. *NeuroImage*. 2011;54(2):1112-1121. doi:10.1016/j.neuroimage.2010.08.068
120. Shams Z, Norris DG, Marques JP. A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. *PLoS ONE*. 2019;14(7). doi:10.1371/journal.pone.0218089
121. Uddin MN, Figley TD, Solar KG, Shatil AS, Figley CR. Comparisons between multi-component myelin water fraction, T1w/T2w ratio, and diffusion tensor imaging measures in healthy human brain structures. *Sci Rep*. 2019;9(1):2500. doi:10.1038/s41598-019-39199-x
122. Connor JR, Menzies SL. Relationship of iron to oligodendrocytes and myelination. *Glia*. 1996;17(2):83-93. doi:10.1002/(SICI)1098-1136(199606)17:2<83::AID-GLIA1>3.0.CO;2-7
123. Ritchie J, Pantazatos SP, French L. Transcriptomic characterization of MRI contrast with focus on the T1-w/T2-w ratio in the cerebral cortex. *NeuroImage*. 2018;174:504-517. doi:10.1016/j.neuroimage.2018.03.027
124. Björnholm L, Nikkinen J, Kiviniemi V, et al. Structural properties of the human corpus callosum: Multimodal assessment and sex differences. *NeuroImage*. 2017;152:108-118. doi:10.1016/j.neuroimage.2017.02.056
125. Alexander DC, Dyrby TB, Nilsson M, Zhang H. Imaging brain microstructure

with diffusion MRI: practicality and applications. *NMR Biomed*. 2019;32(4):e3841. doi:10.1002/nbm.3841
126. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med*. 2005;53(6):1432-1440. doi:10.1002/mrm.20508
127. 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):1000-1016. doi:10.1016/j.neuroimage.2012.03.072
128. Laule C, Leung E, Lis DKB, et al. Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. *Mult Scler Houndmills Basingstoke Engl*. 2006;12(6):747-753. doi:10.1177/1352458506070928
129. Kelly S, Jahanshad N, Zalesky A, et al. Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Mol Psychiatry*. 2018;23(5):1261-1269. doi:10.1038/mp.2017.170
130. Hua K, Zhang J, Wakana S, et al. Tract Probability Maps in Stereotaxic Spaces: Analyses of White Matter Anatomy and Tract-Specific Quantification. *NeuroImage*. 2008;39(1):336-347. doi:10.1016/j.neuroimage.2007.07.053
131. Barth C, Kelly S, Nerland S, et al. In vivo white matter microstructure in adolescents with early-onset psychosis: a multi-site mega-analysis. *Mol Psychiatry*. 2023;28(3):1159-1169. doi:10.1038/s41380-022-01901-3
132. Agartz I, Andersson JL, Skare S. Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study. *Neuroreport*. 2001;12(10):2251-2254. doi:10.1097/00001756-200107200-00041
133. Tønnesen S, Kaufmann T, Doan NT, et al. White matter aberrations and age-related trajectories in patients with schizophrenia and bipolar disorder revealed by diffusion tensor imaging. *Sci Rep*. 2018;8:14129. doi:10.1038/s41598-018-32355-9
134. Tesli N, Westlye LT, Storvestre GB, et al. White matter microstructure in schizophrenia patients with a history of violence. *Eur Arch Psychiatry Clin Neurosci*. 2021;271(4):623-634. doi:10.1007/s00406-019-00988-0
135. Wheeler AL, Voineskos AN. A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. *Front Hum Neurosci*. 2014;8:653. doi:10.3389/fnhum.2014.00653
136. Foong J, Maier M, Barker G, Brocklehurst S, Miller D, Ron M. In vivo investigation of white matter pathology in schizophrenia with magnetisation transfer imaging. *J Neurol Neurosurg Psychiatry*. 2000;68(1):70-74. doi:10.1136/jnnp.68.1.70
137. Palaniyappan L, Al-Radaideh A, Mougin O, Gowland P, Liddle PF. Combined White Matter Imaging Suggests Myelination Defects in Visual Processing Regions in Schizophrenia. *Neuropsychopharmacology*. 2013;38(9):1808-1815. doi:10.1038/npp.2013.80
138. Mandl RCW, Schnack HG, Luigjes J, et al. Tract-based Analysis of Magnetization Transfer Ratio and Diffusion Tensor Imaging of the Frontal and Frontotemporal Connections in Schizophrenia. *Schizophr Bull*. 2010;36(4):778-787. doi:10.1093/schbul/sbn161
139. Palaniyappan L, Al-Radaideh A, Mougin O, Das T, Gowland P, Liddle PF. Aberrant myelination of the cingulum and Schneiderian delusions in

schizophrenia: a 7T magnetization transfer study. *Psychol Med*. 2019;49(11):1890-1896. doi:10.1017/S0033291718002647
140. Vanes LD, Mouchlianitis E, Wood TC, Shergill SS. White matter changes in treatment refractory schizophrenia: Does cognitive control and myelination matter? *NeuroImage Clin*. 2018;18:186-191. doi:10.1016/j.nicl.2018.01.010
141. Vanes LD, Mouchlianitis E, Barry E, Patel K, Wong K, Shergill SS. Cognitive correlates of abnormal myelination in psychosis. *Sci Rep*. 2019;9(1):5162. doi:10.1038/s41598-019-41679-z
142. Kisel AA, Naumova AV, Yarnykh VL. Macromolecular Proton Fraction as a Myelin Biomarker: Principles, Validation, and Applications. *Front Neurosci*. 2022;16. https://www.frontiersin.org/articles/10.3389/fnins.2022.819912. Accessed May 11, 2023.
143. Smirnova LP, Yarnykh VL, Parshukova DA, et al. Global hypomyelination of the brain white and gray matter in schizophrenia: quantitative imaging using macromolecular proton fraction. *Transl Psychiatry*. 2021;11(1):365. doi:10.1038/s41398-021-01475-8
144. Jørgensen KN, Nerland S, Norbom LB, et al. Increased MRI-based cortical grey/white-matter contrast in sensory and motor regions in schizophrenia and bipolar disorder. *Psychol Med*. 2016;46(9):1971-1985. doi:10.1017/S0033291716000593
145. Makowski C, Lewis JD, Lepage C, et al. Structural Associations of Cortical Contrast and Thickness in First Episode Psychosis. *Cereb Cortex*. 2019;29(12):5009-5021. doi:10.1093/cercor/bhz040
146. Iwatani J, Ishida T, Donishi T, et al. Use of T1-weighted/T2-weighted magnetic resonance ratio images to elucidate changes in the schizophrenic brain. *Brain Behav*. 2015;5(10). doi:10.1002/brb3.399
147. Ganzetti M, Wenderoth N, Mantini D. Mapping pathological changes in brain structure by combining T1- and T2-weighted MR imaging data. *Neuroradiology*. 2015;57(9):917-928. doi:10.1007/s00234-015-1550-4
148. Wei W, Zhang Y, Li Y, et al. Depth-dependent abnormal cortical myelination in first-episode treatment-naïve schizophrenia. *Hum Brain Mapp*. 2020;41(10):2782-2793. doi:10.1002/hbm.24977
149. Wei W, Yin Y, Zhang Y, et al. Structural Covariance of Depth-Dependent Intracortical Myelination in the Human Brain and Its Application to Drug-Naïve Schizophrenia: A T1w/T2w MRI Study. *Cereb Cortex*. September 2021:bhab337. doi:10.1093/cercor/bhab337
150. Flynn SW, Lang DJ, Mackay AL, et al. Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins. *Mol Psychiatry*. 2003;8(9):811-820. doi:10.1038/sj.mp.4001337
151. Lang DJM, Yip E, MacKay AL, et al. 48 echo T2 myelin imaging of white matter in first-episode schizophrenia: Evidence for aberrant myelination. *NeuroImage Clin*. 2014;6:408-414. doi:10.1016/j.nicl.2014.10.006
152. Kochunov P, Hong LE, Dennis EL, et al. ENIGMA-DTI: Translating reproducible white matter deficits into personalized vulnerability metrics in cross-diagnostic psychiatric research. *Hum Brain Mapp*. 2020;43(1):194-206. doi:10.1002/hbm.24998
153. Goghari VM, Kusi M, Shakeel MK, et al. Diffusion kurtosis imaging of white matter in bipolar disorder. *Psychiatry Res Neuroimaging*. 2021;317:111341. doi:10.1016/j.pscychresns.2021.111341

154. Benedetti F, Yeh PH, Bellani M, et al. Disruption of White Matter Integrity in Bipolar Depression as a Possible Structural Marker of Illness. *Biol Psychiatry*. 2011;69(4):309-317. doi:10.1016/j.biopsych.2010.07.028
155. Sehmbi M, Rowley CD, Minuzzi L, et al. Association of intracortical myelin and cognitive function in bipolar I disorder. *Acta Psychiatr Scand*. 2018;138(1):62-72. doi:10.1111/acps.12875
156. Sehmbi M, Rowley CD, Minuzzi L, et al. Age-related deficits in intracortical myelination in young adults with bipolar disorder type I. *J Psychiatry Neurosci JPN*. 2019;44(2):79-88. doi:10.1503/jpn.170220
157. Sehmbi M, Suh JS, Rowley CD, et al. Network properties of intracortical myelin associated with psychosocial functioning in bipolar I disorder. *Bipolar Disord*. 2022;24(5):539-548. doi:10.1111/bdi.13181
158. Mahal P, Deep R, Kumaran SS, Khandelwal S. Elevated choline in dorsolateral prefrontal cortex of lithium responders with bipolar I disorder. *Asian J Psychiatry*. 2023;79:103318. doi:10.1016/j.ajp.2022.103318
159. Ćurčić-Blake B, Nanetti L, van der Meer L, et al. Not on speaking terms: hallucinations and structural network disconnectivity in schizophrenia. *Brain Struct Funct*. 2015;220(1):407-418. doi:10.1007/s00429-013-0663-y
160. Ćurčić-Blake B, Ford JM, Hubl D, et al. Interaction of language, auditory and memory brain networks in auditory verbal hallucinations. *Prog Neurobiol*. 2017;148:1-20. doi:10.1016/j.pneurobio.2016.11.002
161. Shao X, Liao Y, Gu L, Chen W, Tang J. The Etiology of Auditory Hallucinations in Schizophrenia: From Multidimensional Levels. *Front Neurosci*. 2021;15. https://www.frontiersin.org/articles/10.3389/fnins.2021.755870. Accessed January 25, 2023.
162. Friederici AD, Gierhan SM. The language network. *Curr Opin Neurobiol*. 2013;23(2):250-254. doi:10.1016/j.conb.2012.10.002
163. Psomiades M, Fonteneau C, Mondino M, et al. Integrity of the arcuate fasciculus in patients with schizophrenia with auditory verbal hallucinations: A DTI-tractography study. *NeuroImage Clin*. 2016;12:970-975. doi:10.1016/j.nicl.2016.04.013
164. De Weijer AD, Neggers SF, Diederen KM, et al. Aberrations in the arcuate fasciculus are associated with auditory verbal hallucinations in psychotic and in non-psychotic individuals. *Hum Brain Mapp*. 2013;34(3):626-634. doi:10.1002/hbm.21463
165. De Weijer AD, Mandl RCW, Diederen KMJ, et al. Microstructural alterations of the arcuate fasciculus in schizophrenia patients with frequent auditory verbal hallucinations. *Schizophr Res*. 2011;130(1-3):68-77. doi:10.1016/j.schres.2011.05.010
166. McCarthy-Jones S, Oestreich LKL, Whitford TJ. Reduced integrity of the left arcuate fasciculus is specifically associated with auditory verbal hallucinations in schizophrenia. *Schizophr Res*. 2015;162(1):1-6. doi:10.1016/j.schres.2014.12.041
167. Knöchel C, Oertel-Knöchel V, Schönmeyer R, et al. Interhemispheric hypoconnectivity in schizophrenia: Fiber integrity and volume differences of the corpus callosum in patients and unaffected relatives. *NeuroImage*. 2012;59(2):926-934. doi:10.1016/j.neuroimage.2011.07.088
168. Mulert C, Kirsch V, Whitford TJ, et al. Hearing voices: a role of interhemispheric auditory connectivity? *World J Biol Psychiatry Off J World Fed Soc Biol Psychiatry*. 2012;13(2):153-158. doi:10.3109/15622975.2011.570789

169. Wigand M, Kubicki M, Hohenberg CC von, et al. Auditory Verbal Hallucinations and the Interhemispheric Auditory Pathway in Chronic Schizophrenia. *World J Biol Psychiatry Off J World Fed Soc Biol Psychiatry*. 2015;16(1):31. doi:10.3109/15622975.2014.948063
170. Hugdahl K, Westerhausen R. Speech processing asymmetry revealed by dichotic listening and functional brain imaging. *Neuropsychologia*. 2016;93(Pt B):466-481. doi:10.1016/j.neuropsychologia.2015.12.011
171. Hugdahl K, Sommer IE. Auditory Verbal Hallucinations in Schizophrenia From a Levels of Explanation Perspective. *Schizophr Bull*. 2018;44(2):234-241. doi:10.1093/schbul/sbx142
172. Hugdahl K. Auditory Hallucinations as Translational Psychiatry: Evidence from Magnetic Resonance Imaging. *Balk Med J*. 2017;34(6):504-513. doi:10.4274/balkanmedj.2017.1226
173. Sato Y, Sakuma A, Ohmuro N, et al. Relationship Between White Matter Microstructure and Hallucination Severity in the Early Stages of Psychosis: A Diffusion Tensor Imaging Study. *Schizophr Bull Open*. 2021;2(1). doi:10.1093/schizbullopen/sgab015
174. Conlon P, Trimble MR, Rogers D. A study of epileptic psychosis using magnetic resonance imaging. *Br J Psychiatry J Ment Sci*. 1990;156:231-235. doi:10.1192/bjp.156.2.231
175. Mørch-Johnsen L, Nesvåg R, Jørgensen KN, et al. Auditory Cortex Characteristics in Schizophrenia: Associations With Auditory Hallucinations. *Schizophr Bull*. 2017;43(1):75-83. doi:10.1093/schbul/sbw130
176. Mørch-Johnsen L, Nerland S, Jørgensen KN, et al. Cortical thickness abnormalities in bipolar disorder patients with a lifetime history of auditory hallucinations. *Bipolar Disord*. 2018;20(7):647-657. doi:10.1111/bdi.12627
177. Natu VS, Gomez J, Barnett M, et al. Apparent thinning of human visual cortex during childhood is associated with myelination. *Proc Natl Acad Sci*. 2019;116(41):20750-20759. doi:10.1073/pnas.1904931116
178. Salat D, Lee S, van der Kouwe A, Greve D, Fischl B, Rosas H. Age-Associated Alterations in Cortical Gray and White Matter Signal Intensity and Gray to White Matter Contrast. *NeuroImage*. 2009;48(1):21-28. doi:10.1016/j.neuroimage.2009.06.074
179. Sagarwala R, Nasrallah HA. The effect of antipsychotic medications on white matter integrity in first-episode drug-naïve patients with psychosis: A review of DTI studies. *Asian J Psychiatry*. 2021;61:102688. doi:10.1016/j.ajp.2021.102688
180. Wang Q, Cheung C, Deng W, et al. White-matter microstructure in previously drug-naive patients with schizophrenia after 6 weeks of treatment. *Psychol Med*. 2013;43(11):2301-2309. doi:10.1017/S0033291713000238
181. Ozcelik-Eroglu E, Ertugrul A, Oguz KK, Has AC, Karahan S, Yazici MK. Effect of clozapine on white matter integrity in patients with schizophrenia: A diffusion tensor imaging study. *Psychiatry Res Neuroimaging*. 2014;223(3):226-235. doi:10.1016/j.pscychresns.2014.06.001
182. Serpa MH, Doshi J, Erus G, et al. State-dependent microstructural white matter changes in drug-naïve patients with first-episode psychosis. *Psychol Med*. 2017;47(15):2613-2627. doi:10.1017/S0033291717001015
183. Zeng B, Ardekani BA, Tang Y, et al. Abnormal white matter microstructure in drug-naive first episode schizophrenia patients before and after eight weeks of antipsychotic treatment. *Schizophr Res*. 2016;172(1):1-8. doi:10.1016/j.schres.2016.01.051

184. Barth C, Lonning V, Gurholt TP, Andreassen OA, Myhre AM, Agartz I. Exploring white matter microstructure and the impact of antipsychotics in adolescent-onset psychosis. *PLOS ONE*. 2020;15(5):e0233684. doi:10.1371/journal.pone.0233684
185. Bartzokis G, Lu PH, Raven EP, et al. Impact on Intracortical Myelination Trajectory of Long Acting Injection Versus Oral Risperidone in First-Episode Schizophrenia. *Schizophr Res*. 2012;140(1-3):122-128. doi:10.1016/j.schres.2012.06.036
186. Raabe FJ, Slapakova L, Rossner MJ, et al. Oligodendrocytes as A New Therapeutic Target in Schizophrenia: From Histopathological Findings to Neuron-Oligodendrocyte Interaction. *Cells*. 2019;8(12):1496. doi:10.3390/cells8121496
187. Gouvêa-Junqueira D, Falvella ACB, Antunes ASLM, et al. Novel Treatment Strategies Targeting Myelin and Oligodendrocyte Dysfunction in Schizophrenia. *Front Psychiatry*. 2020;11:379. doi:10.3389/fpsyt.2020.00379
188. Van Essen DC, Ugurbil K, Auerbach E, et al. The Human Connectome Project: A data acquisition perspective. *NeuroImage*. 2012;62(4):2222-2231. doi:10.1016/j.neuroimage.2012.02.018
189. Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID): I: History, Rationale, and Description. *Arch Gen Psychiatry*. 1992;49(8):624-629. doi:10.1001/archpsyc.1992.01820080032005
190. Jørgensen KN, Nesvåg R, Gunleiksrud S, Raballo A, Jönsson EG, Agartz I. First- and second-generation antipsychotic drug treatment and subcortical brain morphology in schizophrenia. *Eur Arch Psychiatry Clin Neurosci*. 2016;266(5):451-460. doi:10.1007/s00406-015-0650-9
191. Wechsler D. *Wechsler Abbreviated Scale of Intelligence WASI : Manual*. San Antonio: Pearson/PsychCorpl; 2019.
192. Pedersen G, Hagtvet KA, Karterud S. Generalizability studies of the Global Assessment of Functioning–Split version. *Compr Psychiatry*. 2007;48(1):88-94. doi:10.1016/j.comppsych.2006.03.008
193. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*. 1993;88(6):791-804. doi:10.1111/j.1360-0443.1993.tb02093.x
194. Berman AH, Bergman H, Palmstierna T, Schlyter F. Evaluation of the Drug Use Disorders Identification Test (DUDIT) in Criminal Justice and Detoxification Settings and in a Swedish Population Sample. *Eur Addict Res*. 2005;11(1):22-31. doi:10.1159/000081413
195. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophr Bull*. 1987;13(2):261-276. doi:10.1093/schbul/13.2.261
196. Wallwork RS, Fortgang R, Hashimoto R, Weinberger DR, Dickinson D. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. *Schizophr Res*. 2012;137(1):246-250. doi:10.1016/j.schres.2012.01.031
197. Wing JK, Babor T, Brugha T, et al. SCAN: Schedules fonr Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry*. 1990;47(6):589-593. doi:10.1001/archpsyc.1990.01810180089012
198. Andreasen NC. Scale for the assessment of positive symptoms (SAPS).

1984.
199. Andreasen NC. Scale for the assessment of negative symptoms (SANS). 1983.
200. Athinoula A. Martinos Center for Biomedical Imaging. FreeSurfer. FreeSurfer. https://surfer.nmr.mgh.harvard.edu. Accessed May 13, 2023.
201. Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*. 1999;9(2):179-194. doi:10.1006/nimg.1998.0395
202. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*. 1999;9(2):195-207. doi:10.1006/nimg.1998.0396
203. Athinoula A. Martinos Center for Biomedical Imaging. Edits - Free Surfer Wiki. https://surfer.nmr.mgh.harvard.edu/fswiki/Edits. Accessed May 12, 2023.
204. Reinhold J. intensity-normalization. May 2023. https://github.com/jcreinhold/intensity-normalization. Accessed May 12, 2023.
205. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the Impact of Intensity Normalization on MR Image Synthesis. *Proc SPIE-- Int Soc Opt Eng*. 2019;10949. doi:10.1117/12.2513089
206. Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*. 2013;80:105-124. doi:10.1016/j.neuroimage.2013.04.127
207. Shafee R, Buckner RL, Fischl B. Gray matter myelination of 1555 human brains using partial volume corrected MRI images. *NeuroImage*. 2015;105:473-485. doi:10.1016/j.neuroimage.2014.10.054
208. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*. 1998;17(1):87-97. doi:10.1109/42.668698
209. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908
210. Waehnert MD, Dinse J, Weiss M, et al. Anatomically motivated modeling of cortical laminae. *NeuroImage*. 2014;93:210-220. doi:10.1016/j.neuroimage.2013.03.078
211. Shinohara RT, Sweeney EM, Goldsmith J, et al. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage Clin*. 2014;6:9-19. doi:10.1016/j.nicl.2014.08.008
212. Algorithm Descriptions — intensity-normalization 2.2.3 documentation. https://intensity-normalization.readthedocs.io/en/latest/algorithm.html. Accessed May 12, 2023.
213. Maximov II, Alnæs D, Westlye LT. Towards an optimised processing pipeline for diffusion magnetic resonance imaging data: Effects of artefact corrections on diffusion metrics and their age associations in UK Biobank. *Hum Brain Mapp*. 2019;40(14):4146-4162. doi:10.1002/hbm.24691
214. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012;62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015
215. Veraart J, Novikov DS, Christiaens D, Ades-aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *NeuroImage*. 2016;142:394-406. doi:10.1016/j.neuroimage.2016.08.016
216. Kellner E, Dhital B, Kiselev VG, Reisert M. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magn Reson Med*. 2016;76(5):1574-1581. doi:10.1002/mrm.26054

217. 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. doi:10.1016/j.neuroimage.2016.06.058
218. 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. doi:10.1016/j.neuroimage.2015.10.019
219. Warrington S, Bryant KL, Khrapitchev AA, et al. XTRACT - Standardised protocols for automated tractography in the human and macaque brain. *NeuroImage*. 2020;217:116923. doi:10.1016/j.neuroimage.2020.116923
220. Jeurissen B, Descoteaux M, Mori S, Leemans A. Diffusion MRI fiber tractography of the brain. *NMR Biomed*. 2019;32(4):e3785. doi:10.1002/nbm.3785
221. 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. doi:10.1016/j.neuroimage.2021.118502
222. Rheault F, De Benedictis A, Daducci A, et al. Tractostorm: The what, why, and how of tractography dissection reproducibility. *Hum Brain Mapp*. 2020;41(7):1859-1874. doi:10.1002/hbm.24917
223. Eikenes L, Visser E, Vangberg T, Håberg AK. Both brain size and biological sex contribute to variation in white matter microstructure in middle-aged healthy adults. *Hum Brain Mapp*. 2023;44(2):691-709. doi:10.1002/hbm.26093
224. Puonti O, Iglesias JE, Van Leemput K. Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling. *NeuroImage*. 2016;143:235-249. doi:10.1016/j.neuroimage.2016.09.011
225. Nerland S, Stokkan TS, Jørgensen KN, et al. A comparison of intracranial volume estimation methods and their cross‐sectional and longitudinal associations with age. *Hum Brain Mapp*. 2022;43(15):4620-4639. doi:10.1002/hbm.25978
226. Glasser MF, Coalson TS, Harms MP, et al. *Transmit Field Bias Correction of T1w/T2w Myelin Maps*.; 2021:2021.08.08.455570. doi:10.1101/2021.08.08.455570
227. Curtis AT, Gilbert KM, Klassen LM, Gati JS, Menon RS. Slice-by-slice B1+ shimming at 7 T. *Magn Reson Med*. 2012;68(4):1109-1116. doi:10.1002/mrm.23319
228. Weiskopf N, Suckling J, Williams G, et al. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. *Front Neurosci*. 2013;7. doi:10.3389/fnins.2013.00095
229. Leutritz T, Seif M, Helms G, et al. Multiparameter mapping of relaxation (R1, R2*), proton density and magnetization transfer saturation at 3 T: A multicenter dual-vendor reproducibility and repeatability study. *Hum Brain Mapp*. 2020;41(15):4232-4247. doi:https://doi.org/10.1002/hbm.25122
230. van der Weijden CWJ, Biondetti E, Gutmann IW, et al. Quantitative myelin imaging with MRI and PET: an overview of techniques and their validation status. *Brain*. 2023;146(4):1243-1266. doi:10.1093/brain/awac436
231. Cashmore MT, McCann AJ, Wastling SJ, McGrath C, Thornton J, Hall MG. Clinical quantitative MRI and the need for metrology. *Br J Radiol*. 2021;94(1120):20201215. doi:10.1259/bjr.20201215
232. Parent O, Olafson E, Bussy A, et al. High spatial overlap but diverging age-related trajectories of cortical magnetic resonance imaging markers aiming

to represent intracortical myelin and microstructure. *Hum Brain Mapp*. 2023;n/a(n/a). doi:10.1002/hbm.26259
233. Ganzetti M, Wenderoth N, Mantini D. Whole brain myelin mapping using T1- and T2-weighted MR imaging data. *Front Hum Neurosci*. 2014;8. doi:10.3389/fnhum.2014.00671
234. Geoffroy PA, Houenou J, Duhamel A, et al. The arcuate fasciculus in auditory-verbal hallucinations: A meta-analysis of diffusion-tensor-imaging studies. *Schizophr Res*. 2014;159(1):234-237. doi:10.1016/j.schres.2014.07.014
235. Wernicke C. *Der Aphasische Symptomencomplex : Eine Psychologische Studie Auf Anatomischer Basis*. Breslau: Max Cohn & Weigert; 1874.
236. Kong XZ, Group ELW, Francks C. Reproducibility in the absence of selective reporting: An illustration from large-scale brain asymmetry research. *Hum Brain Mapp*. 2022;43(1):244-254. doi:10.1002/hbm.25154
237. Vaher K, Galdi P, Blesa Cabez M, et al. General factors of white matter microstructure from DTI and NODDI in the developing brain. *NeuroImage*. 2022;254:119169. doi:10.1016/j.neuroimage.2022.119169
238. Geeraert BL, Chamberland M, Lebel RM, Lebel C. Multimodal principal component analysis to identify major features of white matter structure and links to reading. *PLOS ONE*. 2020;15(8):e0233244. doi:10.1371/journal.pone.0233244
239. Cox SR, Ritchie SJ, Tucker-Drob EM, et al. Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat Commun*. 2016;7(1):13629. doi:10.1038/ncomms13629
240. Chamberland M, Raven EP, Genc S, et al. Dimensionality reduction of diffusion MRI measures for improved tractometry of the human brain. *NeuroImage*. 2019;200:89-100. doi:10.1016/j.neuroimage.2019.06.020
241. Kanaan RAA, Kim JS, Kaufmann WE, Pearlson GD, Barker GJ, McGuire PK. Diffusion tensor imaging in schizophrenia. *Biol Psychiatry*. 2005;58(12):921-929. doi:10.1016/j.biopsych.2005.05.015
242. Haast RAM, Lau JC, Ivanov D, Menon RS, Uludağ K, Khan AR. Effects of MP2RAGE B1+ sensitivity on inter-site T1 reproducibility and hippocampal morphometry at 7T. *NeuroImage*. 2021;224:117373. doi:10.1016/j.neuroimage.2020.117373
243. Arshad M, Stanley JA, Raz N. Test–retest reliability and concurrent validity of in vivo myelin content indices: Myelin water fraction and calibrated T1w/T2w image ratio. *Hum Brain Mapp*. 2017;38(4):1780-1790. doi:10.1002/hbm.23481
244. Jones DK, Knösche TR, Turner R. White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *NeuroImage*. 2013;73:239-254. doi:10.1016/j.neuroimage.2012.06.081
245. Jeurissen B, Leemans A, Tournier J, Jones DK, Sijbers J. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum Brain Mapp*. 2012;34(11):2747-2766. doi:10.1002/hbm.22099
246. Yano R, Hata J, Abe Y, et al. Quantitative temporal changes in DTI values coupled with histological properties in cuprizone-induced demyelination and remyelination. *Neurochem Int*. 2018;119:151-158. doi:10.1016/j.neuint.2017.10.004
247. Kelm ND, West KL, Carson RP, Gochberg DF, Ess KC, Does MD. Evaluation of diffusion kurtosis imaging in ex vivo hypomyelinated mouse brains. *NeuroImage*. 2016;124:612-626. doi:10.1016/j.neuroimage.2015.09.028
248. Guglielmetti C, Veraart J, Roelant E, et al. Diffusion kurtosis imaging probes

cortical alterations and white matter pathology following cuprizone induced demyelination and spontaneous remyelination. *NeuroImage*. 2016;125:363-377. doi:10.1016/j.neuroimage.2015.10.052
249. Fieremans E, Novikov DS, Jensen JH, Helpern JA. Monte Carlo study of a two-compartment exchange model of diffusion. *NMR Biomed*. 2010;23(7):711-724. doi:10.1002/nbm.1577
250. Fieremans E, Jensen JH, Helpern JA. White matter characterization with diffusional kurtosis imaging. *NeuroImage*. 2011;58(1):177-188. doi:10.1016/j.neuroimage.2011.06.006
251. Jelescu IO, Budde MD. Design and validation of diffusion MRI models of white matter. *Front Phys*. 2017;28:61.
252. Ishida T, Donishi T, Iwatani J, et al. Elucidating the aberrant brain regions in bipolar disorder using T1-weighted/T2-weighted magnetic resonance ratio images. *Psychiatry Res Neuroimaging*. 2017;263:76-84. doi:10.1016/j.pscychresns.2017.03.006
253. Foit NA, Yung S, Lee HM, Bernasconi A, Bernasconi N, Hong SJ. A whole-brain 3D myeloarchitectonic atlas: Mapping the Vogt-Vogt legacy to the cortical surface. *NeuroImage*. 2022;263:119617. doi:10.1016/j.neuroimage.2022.119617
254. Smith SM, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*. 2006;31(4):1487-1505. doi:10.1016/j.neuroimage.2006.02.024

I

NeuroImage 245 (2021) 118709

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Multisite reproducibility and test-retest reliability of the T1w/T2w-ratio: A comparison of processing methods

Stener Nerland[a,b,*], Kjetil N. Jørgensen[a,b], Wibeke Nordhøy[c], Ivan I. Maximov[d,e,f], Robin A.B. Bugge[c], Lars T. Westlye[d,e], Ole A. Andreassen[b,d], Oliver M. Geier[c], Ingrid Agartz[a,b,g]

[a] Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo 0319, Norway
[b] NORMENT, Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[c] Department of Diagnostic Physics, Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway
[d] NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
[e] Department of Psychology, University of Oslo, Oslo, Norway
[f] Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway
[g] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Keywords:
Magnetic resonance imaging
T1w/T2w-ratio
Myelin mapping
Bias field correction
Intensity normalization
Outlier correction
Partial volume correction

ABSTRACT

Background: The ratio of T1-weighted (T1w) and T2-weighted (T2w) magnetic resonance imaging (MRI) images is often used as a proxy measure of cortical myelin. However, the T1w/T2w-ratio is based on signal intensities that are inherently non-quantitative and known to be affected by extrinsic factors. To account for this a variety of processing methods have been proposed, but a systematic evaluation of their efficacy is lacking. Given the dependence of the T1w/T2w-ratio on scanner hardware and T1w and T2w protocols, it is important to ensure that processing pipelines perform well also across different sites.

Methods: We assessed a variety of processing methods for computing cortical T1w/T2w-ratio maps, including correction methods for nonlinear field inhomogeneities, local outliers, and partial volume effects as well as intensity normalisation. These were implemented in 33 processing pipelines which were applied to four test-retest datasets, with a total of 170 pairs of T1w and T2w images acquired on four different MRI scanners. We assessed processing pipelines across datasets in terms of their reproducibility of expected regional distributions of cortical myelin, lateral intensity biases, and test-retest reliability regionally and across the cortex. Regional distributions were compared both qualitatively with histology and quantitatively with two reference datasets, YA-BC and YA-B1+, from the Human Connectome Project.

Results: Reproducibility of raw T1w/T2w-ratio distributions was overall high with the exception of one dataset. For this dataset, Spearman rank correlations increased from 0.27 to 0.70 after N3 bias correction relative to the YA-BC reference and from -0.04 to 0.66 after N4ITK bias correction relative to the YA-B1+ reference. Partial volume and outlier corrections had only marginal effects on the reproducibility of T1w/T2w-ratio maps and test-retest reliability. Before intensity normalisation, we found large coefficients of variation (CVs) and low intraclass correlation coefficients (ICCs), with total whole-cortex CV of 10.13% and whole-cortex ICC of 0.58 for the raw T1w/T2w-ratio. Intensity normalisation with WhiteStripe, RAVEL, and Z-Score improved total whole-cortex CVs to 5.91%, 5.68%, and 5.19% respectively, whereas Z-Score and Least Squares improved whole-cortex ICCs to 0.96 and 0.97 respectively.

Conclusions: In the presence of large intensity nonuniformities, bias field correction is necessary to achieve acceptable correspondence with known distributions of cortical myelin, but it can be detrimental in datasets with less intensity inhomogeneity. Intensity normalisation can improve test-retest reliability and inter-subject comparability. However, both bias field correction and intensity normalisation methods vary greatly in their efficacy and may affect the interpretation of results. The choice of T1w/T2w-ratio processing method must therefore be informed by both scanner and acquisition protocol as well as the given study objective. Our results highlight limitations of the T1w/T2w-ratio, but also suggest concrete ways to enhance its usefulness in future studies.

* Corresponding author.
E-mail address: stener.nerland@medisin.uio.no (S. Nerland).

https://doi.org/10.1016/j.neuroimage.2021.118709.
Received 9 June 2021; Received in revised form 29 October 2021; Accepted 2 November 2021
Available online 27 November 2021.

1. Introduction

Recently, the study of the myelin content of the cerebral cortex has undergone a revival, driven by advances in magnetic resonance imaging (MRI) techniques, a heightened appreciation of the importance of cortical myelin for brain function, and its relevance to a variety of brain disorders. Quantitative relaxometry methods such as R1 (1/T1) and R2* (1/T2*) mapping are considered particularly well-suited for in vivo cortical myelin mapping but are not yet commonly used in clinical studies (Edwards et al., 2018; Uddin et al., 2019). The ratio of T1-weighted (T1w) and T2-weighted (T2w) MRI images has been proposed as a proxy measure of cortical myelin, with the advantage of only requiring conventional MRI sequences. This use of the T1w/T2w-ratio is based on several observations regarding the MRI signal in the cerebral cortex. First, the T1w and T2w image contrast between grey and white matter is mainly determined by myelin (Koenig, 1991; Koenig et al., 1990). Second, the T1 signal can be used to identify and delineate myeloarchitectonic regions in the cerebral cortex (Clark et al., 1992; Eickhoff et al., 2004; Walters et al., 2003). Third, the T2 signal is roughly anticorrelated with myelin, and finally shared field inhomogeneities in the T1w and T2w images are reduced by taking the ratio (Glasser and Van Essen, 2011). In the cerebral cortex, high correlations between R1, a quantitative measure associated with myelin content, and the T1w/T2w-ratio have been reported (Shams et al., 2019), but the microstructural correlates of the T1w/T2w-ratio in white matter are less clear (Uddin et al., 2018).

The use of the T1w/T2w-ratio in neuroimaging research can be placed within three broad categories. First, its regional *intra-subject* distribution has been used to characterise the myeloarchitecture of the cortex, utilising its close spatial correspondence with histological measures of cortical myelin content (Glasser and Van Essen, 2011; Nieuwenhuys and Broere, 2017). For example, the T1w/T2w-ratio is one of the two main structural measures underlying the multi-modal parcellation of the cerebral cortex in Glasser et al. (2016) and its relationship with gradients of cyto- and myeloarchitecture, regional gene expression, and synaptic density are topics of ongoing research (Valk et al., 2020; Wang, 2020). Secondly, the T1w/T2w-ratio is used to characterise *inter-subject* differences in cortical myelin content. In healthy individuals, it has been used for brain age prediction (Rokicki et al., 2020), to characterise cross-sectional age trajectories of cortical myelination (Grydeland et al., 2019), and to study the relationship between cortical myelin and cognitive performance (Grydeland et al., 2013; Norbom et al., 2020; Tzourio-Mazoyer et al., 2019). It has also been studied in clinical conditions with known or hypothesised disturbance of cortical myelin where T1w and T2w images, but not more direct measures of myelin, are commonly acquired. Such conditions include multiple sclerosis (Nakamura et al., 2017; Preziosa et al., 2021; Righart et al., 2017), Huntington's disease (Rowley et al., 2018), Alzheimer's disease (Pelkmans et al., 2019), schizophrenia (Iwatani et al., 2015) and bipolar disorders (Ishida et al., 2017). Finally, the T1w/T2w-ratio has been used in whole-brain analyses (Ganzetti et al., 2014), for example to enhance contrast for brain tissue segmentation (Misaki et al., 2015) and to investigate tissue alterations in multiple sclerosis (Beer et al., 2016; Cooper et al., 2019).

The T1w/T2w-ratio is not, however, a direct measurement of cortical myelin as it is based on intensities expressed in arbitrary units influenced by non-biological variation (Fortin et al., 2016; Shinohara et al., 2014). Residual field inhomogeneities, head placement within the receive coil, image intensity scaling factors, and pulse sequence-specific acquisition parameters are factors that can affect the ratio. Receive coil (B1-) inhomogeneities are routinely corrected for during image acquisition using methods such as Phased array Uniformity Enhancement (PURE) on GE systems, Constant Level AppeaRance (CLEAR) on Philips systems, and Prescan Normalize on Siemens systems. In principle, computing the T1w/T2w-ratio should cancel residual intensity inhomogeneity shared between the T1w and T2w images while enhancing contrast related to myelin content (Glasser and Van Essen, 2011). However, other sources of intensity inhomogeneity, notably bias in the transmit field (B1+), may manifest differently in T1w and T2w images and will therefore not be cancelled in the ratio. It is often assumed that B1+ inhomogeneity profiles are similar across the T1w and T2w images, but this does not hold in general (Glasser and Van Essen, 2011; Sereno et al., 2013). It may therefore be necessary to perform post-hoc bias field correction to minimise residual field inhomogeneities. While dedicated sequences for estimating the B1+ bias field exist, the standard approaches are post-hoc image filters such as the nonparametric nonuniform intensity normalization (N3) algorithm (de Boor, 1972) and its more recent variant N4ITK (Tustison et al., 2010).

Since the T1w/T2w-ratio is formed from two separate image acquisitions, head movement between each pulse sequence may affect the results. Images must therefore be spatially aligned before the ratio is calculated, usually by rigid transformation of the T2w image to the T1w image (Glasser et al., 2013; Glasser and Van Essen, 2011). This step requires robust registration methods and considerable developmental efforts have been directed towards that goal (Greve and Fischl, 2009). However, head movement or poor image coregistration may still cause spurious T1w/T2w-ratio values in isolated voxels. Likewise, the sampling of voxels along reconstructed surfaces is sensitive to errors in the surface placement and this may similarly lead to spurious T1w/T2w-ratio values in isolated surface vertices. For these reasons, outlier correction methods both on the voxel and vertex level have previously been recommended (Glasser and Van Essen, 2011).

A further limitation is the resolution of the T1w and T2w images, typically 1 mm isotropic in clinical studies. Low resolution can lead to greater partial volume effects, whereby intensities within grey matter are contaminated by intensities in cerebrospinal fluid or white matter. To address this issue, Shafee et al. (2015) developed a method for performing partial volume correction (PVC) and found regionally dependent differences between the PVC corrected and uncorrected T1w/T2w-ratio maps, where greater differences were seen in sulci compared to gyri. Depending on the specific application of the T1w/T2w-ratio some of these technical limitations may be more important than others. For example, in vivo parcellation of myeloarchitecture requires reproducible regional distributions of T1w/T2w-values, particularly in boundary regions (Glasser and Van Essen, 2011). In this context, it is less important that T1w/T2w-ratio values are directly comparable across subjects. In contrast, the *precision* of measurements is vital to studies performing *inter-subject* comparisons or investigating associations with biological variables. This is commonly estimated by examining test-retest reliability. A variety of intensity normalisation methods have been proposed to improve the inter-subject comparability of the intensities of T1w and T2w images (Ghassemi et al., 2015; Nyúl and Udupa, 1999; Shah et al., 2011; Shinohara et al., 2014). These methods can be used to correct the T1w/T2w-ratio by applying them to the T1w and T2w images separately before taking the ratio.

In the present study, we implemented several previously proposed methods for computing the T1w/T2w-ratio, including bias field correction and intensity normalisation of the T1w and T2w images prior to taking the ratio, and applied these to four test-retest datasets acquired on four different scanners. These processing pipelines included corrections for field inhomogeneity, partial volume effects and surface outliers, as well as intensity normalisation methods which aim to improve inter-subject comparability of image intensities. We evaluated each pipeline based on two distinct set of criteria. First, correspondence with myelin staining studies is important when using the T1w/T2w-ratio to perform myeloarchitectonic parcellation and depends on good *intra-subject* reproducibility of regional distributions. We therefore assessed the correspondence of T1w/T2w-ratio maps qualitatively with cortical myelin maps from histological staining studies (Nieuwenhuys and Broere, 2017) and quantitatively using rank correlations with in vivo data from Glasser et al. (2016) and from a recent preprint from Glasser et al. (2021). Laterality indices (LIs) were used as a proxy for spatially dependent field bias which may lead to hemispheric

Table 1
T1w and T2w sequences and parameters for each dataset.

	T1w sequence				T2w sequence			
Dataset	NOR-MR750	NOR-Premier	DONDERS	HCP	NOR-MR750	NOR-Premier	DONDERS	HCP
Vendor	General Electric	General Electric	Siemens	Siemens	General Electric	General Electric	Siemens	Siemens
Scanner	Discovery MR750	SIGNA Premier	Magnetom Prisma	Connectom Skyra	Discovery MR750	SIGNA Premier	Magnetom Prisma	Connectom Skyra
Head coil	32-channel	48-channel	32-channel	32-channel	32-channel	48-channel	32-channel	32-channel
Sequence	BRAVO	MPRAGE	MPRAGE	MPRAGE	CUBE	CUBE	SPACE	SPACE
Resolution	1 mm iso	0.8 mm iso	0.8 mm iso	0.7 mm iso	1 mm iso	0.8 mm iso	0.8 mm iso	0.7 mm iso
TA (mm:ss)	04:43	06:22	06:02	07:40	04:23	05:21	06:19	08:24
TR (ms)	8.16	2356	2200	2400	2500	3202	3200	3200
TE (ms)	3.18	3.13	2.64	2.14	71.68	90.6	569	565
TI (ms)	450	950	1100	1000	–	–	–	–
FA (deg)	12	8	11	8	90	Variable	Variable	Variable
BW (Hz/px)	244	238	170	210	488	555	579	744
iPAT	2	2	3	2	2 × 2	2 × 2	3	2
B1- calibration	PURE	PURE	Prescan Normalise	No	PURE	PURE	Prescan Normalise	No

intensity biases. Second, acceptable test-retest reliability is important when using the T1w/T2w-ratio for *inter-subject* comparisons. We therefore evaluated the test-retest reliability of the T1w/T2w-ratio via coefficients of variation (CVs) and intraclass correlation coefficients (ICCs). This gives an estimate of the precision which is an important constraint on the detectible effects when using the measure (Matheson, 2019).

2. Materials and methods

2.1. Participants

Participants were included from four test-retest datasets acquired on two 3T GE scanners (GE Medical Systems, Milwaukee, USA) and two 3T Siemens scanners (Siemens, Erlangen, Germany). There was an overlap of eight participants for the NOR-MR750 and NOR-Premier datasets. For these participants, the scan-rescan interval between the first session on in the NOR-MR750 and the first session on the NOR-Premier dataset ranged from 9.8 to 10 months. Scanner model, sequence types and acquisition parameters for each dataset is given in Table 1.

2.1.1. Dataset 1 - The Norwegian centre for mental disorders research - general electric discovery MR750 (NOR-MR750)

This dataset was collected as part of a reliability study conducted at the Norwegian Centre for Mental Disorders Research (NORMENT). Nine healthy participants (mean age = 35.76 years; range = [26.31–59.70]; 55% male) were recruited internally and scanned on a 3T GE Discovery MR750 equipped with a 32-channel receive coil in two sessions with a two-week interval between sessions. This interval was included to account for the influence of intra-subject variability related to day-to-day variation. Each participant was scanned at the same time of day across sessions. In each session two pairs of T1w and T2w MRI images were acquired, with repositioning between each pair, resulting in four T1w and four T2w images for each participant per session.

2.1.2. Dataset 2 - The Norwegian centre for mental disorders research - general electric signa premier (NOR-Premier)

This dataset was acquired as a continuation of the NOR-MR750 test-retest dataset after a major scanner upgrade to 3T GE SIGNA Premier equipped with a 48-channel head coil on the same magnet. The dataset consisted of nine healthy participants (mean age = 35.80 years; range = [25.24–60.52]; 55% male) of which eight participants overlapped with the NOR-MR750 dataset. As with the NOR-MR750 dataset, a pair of test-retest acquisitions were made in two separate sessions with a two week interval between sessions. The interval between the first session of the NOR-MR750 dataset and the first session of the NOR-Premier dataset was 9.6 months on average.

2.1.3. Dataset 3 - the Donders centre for cognitive neuroimaging - siemens magnetom prisma (Donders)

Seventeen healthy participants (mean age = 24.7 ± 2.8 years) were enroled in a test-retest study at the Donders Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, Netherlands (https://www.ru.nl/donders/). Participants were scanned on a 3T Siemens Magnetom Prisma equipped with a 32-channel head coil. Each participant was scanned twice with repositioning between each pair of T1w and T2w acquisitions and randomised head fixation either with cushions only or cushions and a chin-rest.

2.1.4. Dataset 4 - human connectome project - siemens connectom Skyra - (HCP)

A subset of 34 participants (mean age = 30.70 years; range = [22–35]; 29% male) with a scan-rescan interval of less than six months (mean = 3.85 months; range = [2–5]) were selected from test-retest acquisitions included as part of the 1200 Subjects release (S1200 Release, February 2017) by the WU-Minn Human Connectome Project (HCP) (Van Essen et al., 2013). Participants were recruited as part of the HCP Young Adult study and were scanned twice on a customised 3T Siemens Connectom Skyra scanner using a standard 32-channel receive head coil.

2.1.5. Ethics statement

The study was carried out in accordance with the Helsinki Declaration. Participants provided written informed consent at each site and agreed to participation in scientific studies. Each site had ethical approval from their local ethics committees. Data was handled according to guidelines set forth by the Norwegian Data Protection Authority in compliance with GDPR regulations.

2.2. MRI preprocessing

All scans were processed in FreeSurfer (version 7.1.0; https://surfer.nmr.mgh.harvard.edu/) with both T1w and T2w images used as input and the *-T2pial* flag enabled. Briefly, FreeSurfer segments cortical and subcortical tissue compartments, and creates surface meshes representing the boundaries between grey and white matter (WM) and grey matter (GM) and non-brain. For more details on the FreeSurfer processing steps see Fischl (2012). For the HCP dataset, the PreFreeSurfer pipeline had already been used to create an undistorted structural image for the subject in native space, perform the initial brain extraction, coregister T1w and T2w images, perform bias field correction using the square of the T1w and T2w images, and transform the data to MNI space. The PreFreeSurfer pipeline was applied to the HCP dataset for all pipelines due to technical considerations particular to this dataset, including the application of gradient nonlinearity distortion in images acquired with oblique slices relative to the scanner

coordinate system and the lack of on-scanner B1- correction, following the recommendations in Glasser et al. (2013). All images and surface reconstructions were quality controlled by trained research assistants and manual editing was performed following standard FreeSurfer guidelines in the event of surface reconstruction errors. A total number of 170 T1w/T2w pairs were processed, composed of nine participants in the NOR-MR750 dataset and nine participants in the NOR-Premier dataset each scanned four times (scan-rescan in two sessions), 15 participants in the Donders dataset (after three participants were excluded) and 34 participants in the HCP dataset each scanned twice (scan-rescan on the same day and across two sessions respectively).

For the HCP Minimal Processing Pipeline (HCP-MPP), FreeSurfer v6.0.0 and the PreFreeSurfer, FreeSurfer and PostFreeSurfer pipelines were used. The Connectome Workbench was used to map each myelin map computed with these pipelines, as well as the previously published HCP Multi-Modal Parcellation (HCP-MMP; Glasser et al., 2016) scheme, from the *FS_LR* surface to the standard FreeSurfer *fsaverage* surfaces as described on the HCP Wiki (https://wiki.humanconnectome.org/download/attachments/63078513/Resampling-FreeSurfer-HCP.pdf).

For all pipelines, T2w images were rigidly coregistered to T1w images using *bbregister* in FreeSurfer with FSL initialisation (Greve and Fischl, 2009). The resolution of the datasets included 0.7 mm for the HCP dataset, 0.8 mm for the Donders and NOR-Premier dataset, and 1 mm isotropic voxel resolutions for the NOR-MR750 dataset. In order to facilitate the comparison between the different datasets, and to ensure that the results would be generalisable to the more commonly used 1 mm isotropic voxel resolution in clinical settings, we conformed all images to 1 mm prior to FreeSurfer and T1w/T2w-ratio processing. Follow-up analyses were conducted to assess the effect of downsampling (Section 2.4.8). Throughout, *fslmaths* and *fslstats* from FSL (v6.0.4; https://fsl.fmrib.ox.ac.uk/fsl/) were used for image operations and preprocessing was done on the TSD (Service for Sensitive Data) computing cluster at the University of Oslo.

2.3. Calculation of T1w/T2w-ratio maps

We first implemented corrections for partial voluming, surface outliers and bias field in nine different pipelines using local and non-linear correction methods for computing the T1w/T2w-ratio. Here, the goal was to assess pipelines on their ability to reproduce the expected regional distributions of the *intra-subject* T1w/T2w-ratio. These processing pipelines included the raw ratio ($T1w/T2w_{Raw}$), N3 or N4ITK bias field correction only ($T1w/T2w_{N3}$ & $T1w/T2w_{N4}$), partial volume correction only ($T1w/T2w_{PVC}$), surface outlier only ($T1w/T2w_{OC}$), N3 or N4ITK bias field correction with partial volume and outlier correction ($T1w/T2w_{N3\text{-}All}$ & $T1w/T2w_{N4\text{-}All}$), and finally the HCP Minimal Processing Pipeline with ($T1w/T2w_{HCP\text{-}BC}$) and without ($T1w/T2w_{HCP\text{-}Raw}$) template-based bias correction (Section 2.3.6). We qualitatively assessed the correspondence of T1w/T2w-ratio maps with histological myelin maps (Section 2.4.2) and computed Spearman rank correlations with the YA-BC and the YA-B1+ reference datasets (Sections 2.3.7 and 2.4.3), and used the Laterality Index (Section 2.4.6) as a measure of the removal of lateral intensity bias.

Secondly, in order to assess and improve the test-retest reliability of the T1w/T2w-ratio, i.e. the consistency between repeated measurements, we evaluated T1w/T2w-ratio maps after applying eight different intensity normalisation methods (Section 2.3.5). These intensity normalisation methods were applied to the $T1w/T2w_{Raw,}$ $T1w/T2w_{N3\text{-}All}$, and the $T1w/T2w_{N4\text{-}All}$ pipelines. The main outcomes of interest here were coefficients of variation (Section 2.4.4 and 2.4.7), a standardised measure of dispersion, and intraclass correlation coefficients (Section 2.4.5 and 2.4.7) as a measure of the agreement between groups of measurements. It was assumed that true intra-subject scan-rescan differences are small, and smaller for scan-rescan pairs closer in time, relative to inter-subject variation which is ideally still present after intensity normalisation. In total, we evaluated $9 + 8 \times 3 = 33$ separate pipelines. An overview of the processing pipelines implemented in this study is shown in Fig. 1.

2.3.1. Surface projection

For all processing methods, T1w/T2w-ratio intensities in grey matter were sampled along equivolumetric surfaces at distances of 10% to 80% of cortical thickness from the white-grey surface created with the Surface Tools toolbox (https://github.com/kwagstyl/surface_tools). This toolbox reconstructs intermediary surfaces with improved correspondence with cortical laminae compared to standard sampling methods (Waehnert et al., 2014). The values of vertices in the medial wall were set to zero for all figures and analyses.

2.3.2. Bias field correction

For bias field correction, we used both N3 bias correction (Sled et al., 1998) and the more recent N4ITK algorithm (Tustison et al., 2010). Both bias field correction methods were used with a shrink factor of 1 and masked with the whole-brain segmentation created in the standard FreeSurfer pipeline.

2.3.3. Partial volume correction

Partial volume correction was performed as described in Shafee (2015) using the implementation included in FreeSurfer. This method first calculates the proportion of voxels in an upsampled MR image falling within the cortical ribbon, i.e. between the reconstructed pial and white-grey surfaces, and then inverts systems of linear equations to recover the “true” intensity values adjusted for partial volume effects.

2.3.4. Surface outlier correction

Surface-based outlier correction was performed as in Glasser et al. (2011). This algorithm calculates, for each vertex, the mean and standard deviation (SD) of normal-appearing vertex values (only values within 3 SDs of the total mean T1w/T2w-ratio) in a neighbourhood of 10 steps along the surface and replaces values, in vertices exceeding this mean by more than 2 SDs, with a Gaussian-weighted mean of its neighbours. The algorithm was implemented in MATLAB (The Mathworks, Inc., Massachusetts, USA) using the *geodesic* toolbox (Wang, 2021) for computing distances between vertices.

2.3.5. Intensity normalisation

To assess the effect of intensity normalisation, we applied eight different normalisation methods to T1w/T2w-ratio maps from the $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ processing methods which were compared to the unnormalised $T1w/T2w_{Raw}$, and $T1w/T2w_{HCP\text{-}Raw}$ processing methods. All intensity normalisation methods were based on the implementations in the Intensity Normalization toolbox (https://github.com/jcreinhold/intensity-normalization; Reinhold et al., 2019) with more detailed documentation available here: https://intensity-normalization.readthedocs.io/. Intensity normalisation was performed on T1w and T2w images separately.

The eight normalisation methods used in this study were: 1) Z-score (ZS), which subtracts the mean intensity of the skull-stripped image and divides by the standard deviation of the same region, 2) Fuzzy C-Means (FCM; Chen and Giger, 2004), which determines a three-class segmentation using fuzzy C-means clustering, and then normalises the image to a standard value using the mean of a specified tissue class, in our case GM, 3) Gaussian Mixture Model (GMM), which fits a gaussian mixture to the skull-stripped image and translates the WM mean to a standard value, 4) Kernel Density Estimation (KDE), which uses kernel density estimation to find the peak of WM intensities and translates this peak to a standard value, 5) Nyúl and Udupa (N&U; Nyúl and Udupa, 1999), which uses piecewise linear histogram matching to estimate a transformation that standardises the histograms of a population of images, 6) WhiteStripe (WS; Shinohara et al., 2014), which identifies normal-appearing WM voxels, subtracts the mean of this region from the image and finally divides the image by its standard deviation, 7) Removal of Artificial

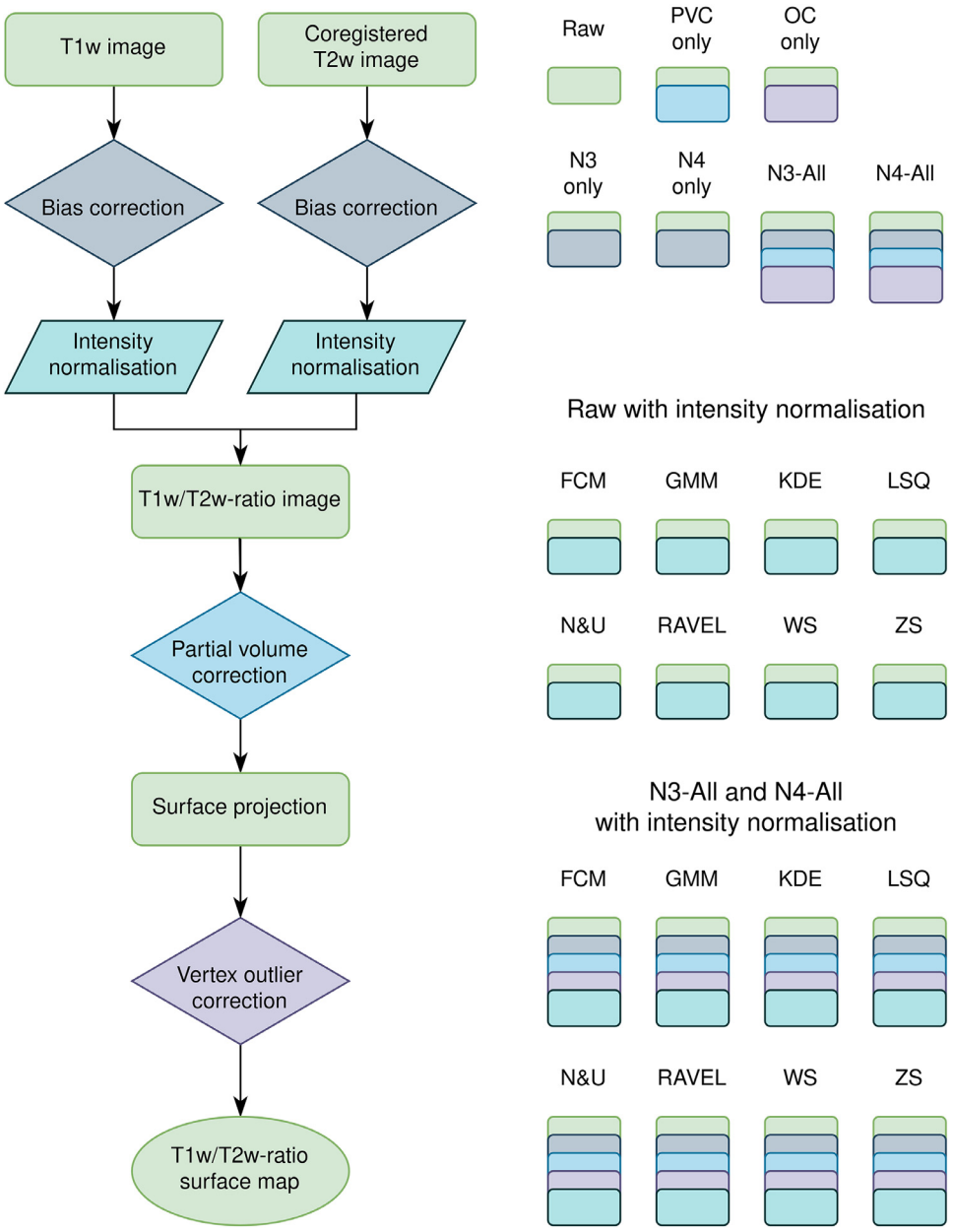


Fig. 1. Overview of the processing methods implemented in this study. Each box and colour on the left corresponds to a step in the pipeline. Collections of boxes on the right correspond to a pipeline. On the top are the pipelines $T1w/T2w_{Raw}$, $T1w/T2w_{PVC}$, $T1w/T2w_{OC}$, $T1w/T2w_{N3}$, $T1w/T2w_{N4}$, $T1w/T2w_{N3\text{-}All}$ and $T1w/T2w_{N4\text{-}All}$ which were primarily assessed on their ability to reproduce the expected T1w/T2w-ratio distributions. On the bottom are the pipelines $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ with each intensity normalisation method. $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ were computed in the separate HCP Minimal Processing Pipeline and are not depicted.

Voxel Effect by Linear regression (RAVEL; Fortin et al., 2016), which first uses WhiteStripe and then estimates the unwanted non-biological variation from a control region which is then regressed out for each voxel in the image, and finally 8) Least Squares (LSQ), which minimises the squared distance between CSF, GM and WM for a population of images. Of these intensity normalisation methods, N&U, RAVEL, and LSQ are population-based methods that are applied to a set of images, while the other methods are used on each image separately.

2.3.6. HCP-MPP T1w/T2w-ratio maps

The T1w/T2w-ratio maps of the HCP-MPP are calculated in the PostFreeSurfer pipeline. Briefly, in this pipeline T1w/T2w-ratio values are projected to a midthickness surface as described in Glasser & Van Essen (2011) and individual surfaces are mapped to the population-average Conte69 surface. The vertex-wise difference between individual $T1w/T2w_{HCP\text{-}Raw}$ maps and the average Conte69 map is then smoothed with a large kernel and subtracted from the $T1w/T2w_{HCP\text{-}Raw}$ maps to create the $T1w/T2w_{HCP\text{-}BC}$ maps. The goal of this procedure is to replace spurious low frequency variation in individual maps with values from the more robust group map. The resulting $T1w/T2w_{HCP\text{-}BC}$ maps therefore closely match the Conte69 average, having undergone both a template-based bias field correction as well as a nonlinear intensity normalisation. This is in contrast to the linear intensity normalisation methods described in Section 2.3.5 and results in a regional shift towards the Conte69 template precluding direct inter-subject comparisons of T1w/T2w-ratio values. See Glasser & Van Essen (2011) and Glasser et al. (2013) for more details.

2.3.7. Reference T1w/T2w-ratio datasets

As reference T1w/T2w-ratio datasets, we used two publicly available T1w/T2w-ratio datasets from the HCP Young Adult (HCP-YA) project acquired using the HCP MRI protocol (Uğurbil et al., 2013; Van Essen et al., 2013). The first reference dataset, YA-BC, consists of T1w/T2w-ratio maps from 210 participants (aged 22–35 years) and is available through the HCP-500 data release. These maps were computed using the $T1w/T2w_{HCP\text{-}BC}$ pipeline and were used, together with thickness and fMRI, in the creation of the HCP-MMP parcellation in Glasser et al. (2016). The dataset is available online at:

https://balsa.wustl.edu/WDpX. The second reference dataset, YA-B1+, consists of T1w/T2w-ratio maps from 1042 participants (aged 22–35 years) and is available through the HCP-1200 data release. These maps were computed with the $T1w/T2w_{HCP\text{-}Raw}$ pipeline with the addition of a new correction procedure, where a separate actual flip angle (AFI) sequence was used to adjust the T1w/T2w-ratio maps for residual B1+ field effects instead of the template-adjustment method used in the $T1w/T2w_{HCP\text{-}BC}$. For details on this procedure see the preprint by Glasser et al. (2021).

For both reference datasets, mean T1w/T2w-ratio values for each region of interest (ROI) in the HCP-MMP were extracted directly from the native *FS_LR* surface using the Connectome Workbench. For the YA-B1+ dataset, we extracted data corresponding to the group correction method described in the recent preprint by Glasser et al. (2021). This dataset is available at: https://balsa.wustl.edu/study/show/mDBP0.

2.4. Data analysis

Analyses aimed to answer two main questions: 1) What is the reproducibility of regional T1w/T2w-ratio distributions compared to the expected distribution from the literature on cortical myelin and can local and nonlinear correction factors improve it? 2) What is the test-retest reliability of the T1w/T2w-ratio and can intensity normalisation methods be used to improve it?

2.4.1. Data quality control

All images were visually inspected prior to analysis to ensure acceptable image quality without large image artefacts such as severe ringing due to movement, misregistration between the T1w and T2w images, or inconsistencies in image intensity scale or acquisition parameters. This resulted in the exclusion of the three first subjects in the Donders dataset due to one missing T2w image, one image with ringing artefacts, and one image with an intensity range suggestive of acquisition or reconstruction differences compared to the other images. No other data was excluded.

2.4.2. Qualitative assessment of cortical maps

To qualitatively assess the regional correspondence with known cortical myelin distributions, we created mean T1w/T2w-ratio maps for each of the non-intensity normalised processing methods. The correspondence of these mean maps was assessed visually against cortical myelin maps based on histological staining studies, which we considered to be the gold standard. These maps are based on extensive work spanning the period from 1910 to 1970, later composed to form a map of the human cortex (Nieuwenhuys and Broere, 2017). We also compared the mean T1w/T2w-ratio maps to previously published T1w/T2w-ratio maps of the YA-BC dataset (Glasser et al., 2016) and of the YA-B1+ where a new adjustment method to account for B1+ field effects has been applied as described in the recent preprint by Glasser et al. (2021). Based on this literature, we located regions known to exhibit high, intermediate, or low cortical myelination. We then assessed the T1w/T2w-ratio maps and reported where the maps diverged from expectations. Briefly, we expected high T1w/T2w-ratio values in primary sensory and motor regions as well as early association cortices, and lower T1w/T2w-ratio values in higher association cortices and in frontal regions.

The regions that were expected to exhibit high levels of cortical myelin were: 1) the motor-somatosensory strip in the central sulcus on the lateral side and extending into the medial 'notch' in the paracentral lobule, 2) the visual cortex in the occipital lobe extending into the temporal and parietal lobes, and 3) early auditory areas in the Sylvian fissure and the upper segment of the temporal lobe. Regions where we expected intermediate cortical myelin levels were: 1) the inferior parietal cortex, and 2) most of the temporal, prefrontal, cingulate and medial and superior parietal cortices. The lowest myelin levels were expected in: 1) the anterior insula, 2) temporal pole, 3) medial prefrontal cortex, and 4) portions of the anterior cingulate cortex. Consensus was reached by comparing notes on each T1w/T2w-ratio map. For assessments of the gradients, we examined if the transitions from high-to-intermediate, high-to-low and intermediate-to-low myelination matched expectations as described above. See Nieuwenhuys & Broere (2017) and Glasser & Van Essen (2011) for further information on the distribution of myelin in the human cerebral cortex.

2.4.3. ROI-wise correlations with reference datasets

To quantify the correspondence with expected myelin distributions, we extracted, for each dataset and processing method, mean T1w/T2w-ratio values in 176 ROIs of the HCP Multi-Modal Parcellation (HCP-MMP), where 4 ROIs (H_ROI, EC_ROI, PreS_ROI and RSC_ROI) were excluded for each hemisphere due to their overlapping with the medial wall yielding spurious T1w/T2w-ratio values. The mean values were calculated across all the T1w/T2w-ratio maps in that dataset, i.e. each test-retest pair in the HCP and Donders datasets and test-retest pair for each of the two sessions in the NOR-MR750 and NOR-Premier datasets. They were then used to compute Spearman rank correlations, using the *cor.test* function from the *stats* package in R, with mean T1w/T2w-ratio values from two reference T1w/T2w-ratio datasets, the YA-BC and YA-B1+ (Section 2.3.7). These datasets were used as references given their acquisition parameters optimised for T1w/T2w-ratio mapping and high correspondence with histology.

2.4.4. Coefficients of variation

To quantify test-retest reliability, we calculated percentage CVs by first calculating vertex-wise CVs which were combined for each participant using the root-mean-square (RMS) formula in each HCP-MMP ROI and for the whole cortex. The ROI and whole-cortex CVs were then combined across individuals to form summary ROI and whole-cortex CVs for each processing method and pipeline. To compute the vertex-wise CVs for each participant we used the formula

$$CV_{Vertex,i,j} = \frac{sd(x_{i,j})}{mean(x_{i,j})}$$

where $x_{i,j}$ is the set of T1w/T2w-ratio values for the i^{th} subject in the j^{th} vertex. For each subject, these were combined for each ROI in the HCP-MMP atlas and for the whole cortical surface to form individual regional coefficients of variation ($CV_{ROI,i}$) and whole-cortex coefficients of variation ($CV_{Cortex,i}$) using the RMS formula as follows

$$CV_{ROI,i} = \sqrt{\sum_{j=1}^{N_{ROI}} \frac{CV_{Vertex,i,j}^2}{N_{ROI}}}$$

where the sum is over the N_{ROI} vertices in each ROI and the whole cortex. To combine ROI and whole-cortex coefficients of variation across all participants, we used the subject-wise CVs calculated above in the RMS formula

$$CV_{ROI} = \sqrt{\sum_{i=1}^{N_{Subjects}} \frac{CV_{ROI,i}^2}{N_{Subjects}}}$$

where the sum is taken over all the subjects in each dataset. Here the RMS is used rather than the arithmetic mean, since the mean is a biased estimator and will tend to underestimate the true test-retest variability (Glüer et al., 1995). Lower CVs indicate higher measurement precision.

2.4.5. Intraclass correlation coefficients

To assess global test-retest variability of the T1w/T2w-ratio distributions, we calculated intraclass correlation coefficients (ICCs) of the median of the T1w/T2w-ratio values (ICC_{Median}), using the *icc* function from the *irr* package in R. We used the two-way mixed single score ICC, denoted ICC(A,1) in the classification by McGraw & Wong (1996). The median was used as a more robust descriptor of the global T1w/T2w-ratio distribution. Higher ICCs represent better absolute agreement between test and retest.

2.4.6. Laterality indices

To assess how correction methods, in particular bias field correction, affected lateral bias in T1w/T2w-ratio values, we computed average interhemispheric T1w/T2w-ratio differences using the percentage laterality index (LI) for each of the 180 regions of the HCP-MMP atlas with the mean laterality index

$$LI = \frac{200}{180}\sum_{j=1}^{180}\frac{ROI_{j,lh} - ROI_{j,rh}}{ROI_{j,lh} + ROI_{j,rh}}$$

Assuming that inter-hemispheric distributions of cortical myelin is not systematically skewed, higher laterality indices indicate the presence of field bias and successful bias field correction should result in a lower mean laterality index.

2.4.7. Visualisation of results

To visualise regional distributions of CV_{ROI} and ICC_{ROI} we created surface maps based on the HCP-MMP atlas using the *fsbrain* toolbox in R (https://github.com/dfsp-spirit/fsbrain). For comparison of the test-retest reliability and laterality indices of each processing method and visualisation of outlier subjects, we created box plots of the whole-cortex CVs, median ROI-wise ICCs and mean laterality indices across ROIs for each participant grouped by processing method and dataset. To visualise the test-retest reliability and intra-subject variation compared to inter-subject variation we created ridgeline plots of the T1w/T2w-ratio distributions for each processing method using the *ggridges* toolbox in R (https://wilkelab.org/ggridges/) which are presented in the Supplementary Materials.

2.4.8. Follow-up analyses

Three of the datasets included in this study were comprised of high-resolution images that were downsampled to 1 mm isotropic to ensure that results were generalisable to commonly available datasets. Past studies have suggested that the use of high-resolution data is beneficial for T1w/T2w-ratio mapping (Glasser and Van Essen, 2011). We therefore performed follow-up analyses for the NOR-Premier and the HCP datasets where we assessed the effect of downsampling by running all analyses described above for $T1w/T2w_{Raw\text{-}Hires}$, $T1w/T2w_{HCP\text{-}Raw\text{-}Hires}$, and T1w/T2wHCP-BC-Hires where we used high-resolution data. The Donders dataset was not included in these analyses, due to the incompatibility of the *-cw256* flag, which was necessary in this dataset due to defacing, and the *-hires* flag of the high-resolution *recon-all* pipeline.

In our main analyses, we transformed output from the $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ pipelines from the native *FS_LR* surface of the HCP-MPP to the native surface of FreeSurfer, *fsaverage*. This transformation could be a source of error, and we therefore tested its effects by extracting data directly from *FS_LR* space for the $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ pipelines using Connectome Workbench and computing ROI-wise correlations with the reference datasets.

3. Results

3.1. Qualitative assessment of T1w/T2w-ratio maps

Mean maps for $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, $T1w/T2w_{N4\text{-}All}$, $T1w/T2w_{HCP\text{-}Raw}$, and $T1w/T2w_{HCP\text{-}BC}$ for each dataset are presented in Fig. 2. Intensity normalisation and corrections for outliers or partial volume effects did not appreciably influence the T1w/T2w-ratio maps and these are therefore not depicted or discussed in the qualitative assessment. See Supplementary Figures 1–5 for ridgeline plots of the T1w/T2w-ratio distributions for each processing method.

3.1.1. NOR-MR750

As expected from histology, high $T1w/T2w_{Raw}$ values were seen in the lateral motor cortex and in the auditory cortex, but the visual cortex had relatively low values despite this region being known to be highly myelinated. Some frontal medial regions had relatively high T1w/T2w-ratio values despite these regions being known to be lightly myelinated. N3 bias correction improved the correspondence with cortical myelin maps, with higher values in the medial occipital lobe but with no clearly delineated notch in the paracentral lobule. N4ITK bias correction led to an improved correspondence with histology on the medial side, but parietal and superior frontal regions with spuriously high values were observed with unexpectedly high values around Brodmann areas 44 and 45. $T1w/T2w_{HCP\text{-}Raw}$ was similar to $T1w/T2w_{Raw}$, with additional smoothing inherent to the HCP Minimal Processing Pipeline. The best correspondence with cortical myelin maps was seen with $T1w/T2w_{HCP\text{-}BC}$, where high values were found in the auditory, motor and visual cortices as expected and a clearly delineated notch was present in the paracentral lobule.

3.1.2. NOR-Premier

$T1w/T2w_{Raw}$ closely matched histological cortical myelin maps with high values in the auditory, motor and visual cortices and low values in frontal, temporal and parietal regions. N3 bias correction only had a marginal effect on the T1w/T2w-ratio maps, but N4ITK bias correction introduced spuriously high values especially in frontal and parietal regions where cortical myelin content is known to be low. $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ gave similar results to $T1w/T2w_{Raw}$ and $T1w/T2w_{N3\text{-}All}$ with some additional smoothing.

3.1.3. Donders

$T1w/T2w_{Raw}$ values were high in the auditory, visual and motor cortices as expected, but lightly myelinated frontal and parietal regions contained spuriously high T1w/T2w-ratio values. N3 bias correction improved the visual correspondence with histology, but N4ITK bias correction seemed to further exacerbate the poor correspondence in regions with spuriously high T1w/T2w-ratio values. $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ were similar, but $T1w/T2w_{HCP\text{-}BC}$ had a better contrast between high and low myelinated regions, with relatively lower T1w/T2w-ratio values in frontal parietal regions in particular, and thus gave the best results visually.

3.1.4. HCP

Regional distributions of $T1w/T2w_{Raw}$ closely matched histology, with the exception of some frontal regions with spuriously high values especially in the left hemisphere. N3 bias correction had only minor effects on the T1w/T2w-ratio, but N4ITK bias correction seemed to reduce the presence of spuriously high values in frontal regions. $T1w/T2w_{HCP\text{-}Raw}$ showed excellent correspondence with the expected cortical myelin distribution, but unexpectedly high values were seen in a frontal parietal region. This region was attenuated in $T1w/T2w_{HCP\text{-}BC}$, which gave the best results.

3.2. ROI-wise correlations with reference datasets

See Table 2 for Spearman rank correlations between the YA-BC and YA-B1+ reference datasets for each of the non-intensity normalised processing methods, and Tables 3–5 for Spearman rank correlations after intensity normalisation applied to $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ respectively.

3.2.1. NOR-MR750

We found low correspondence with the YA-BC reference for $T1w/T2w_{Raw}$ [$\rho = 0.27$], $T1w/T2w_{OC}$ [$\rho = 0.27$], $T1w/T2w_{PVC}$ [$\rho = 0.26$], and $T1w/T2w_{HCP\text{-}Raw}$ [$\rho = 0.23$]. This was greatly improved for $T1w/T2w_{N3}$ [$\rho = 0.70$] and $T1w/T2w_{N3\text{-}All}$ [$\rho = 0.70$], and moderately improved for $T1w/T2w_{N4}$ [$\rho = 0.53$] and $T1w/T2w_{N4\text{-}All}$ [$\rho = 0.57$]. The best results were seen for $T1w/T2w_{HCP\text{-}BC}$ [$\rho = 0.92$].

We found very low correspondence with the YA-B1+ reference, for $T1w/T2w_{Raw}$ [$\rho = -0.05$], $T1w/T2w_{OC}$ [$\rho = -0.04$], $T1w/T2w_{PVC}$ [ρ

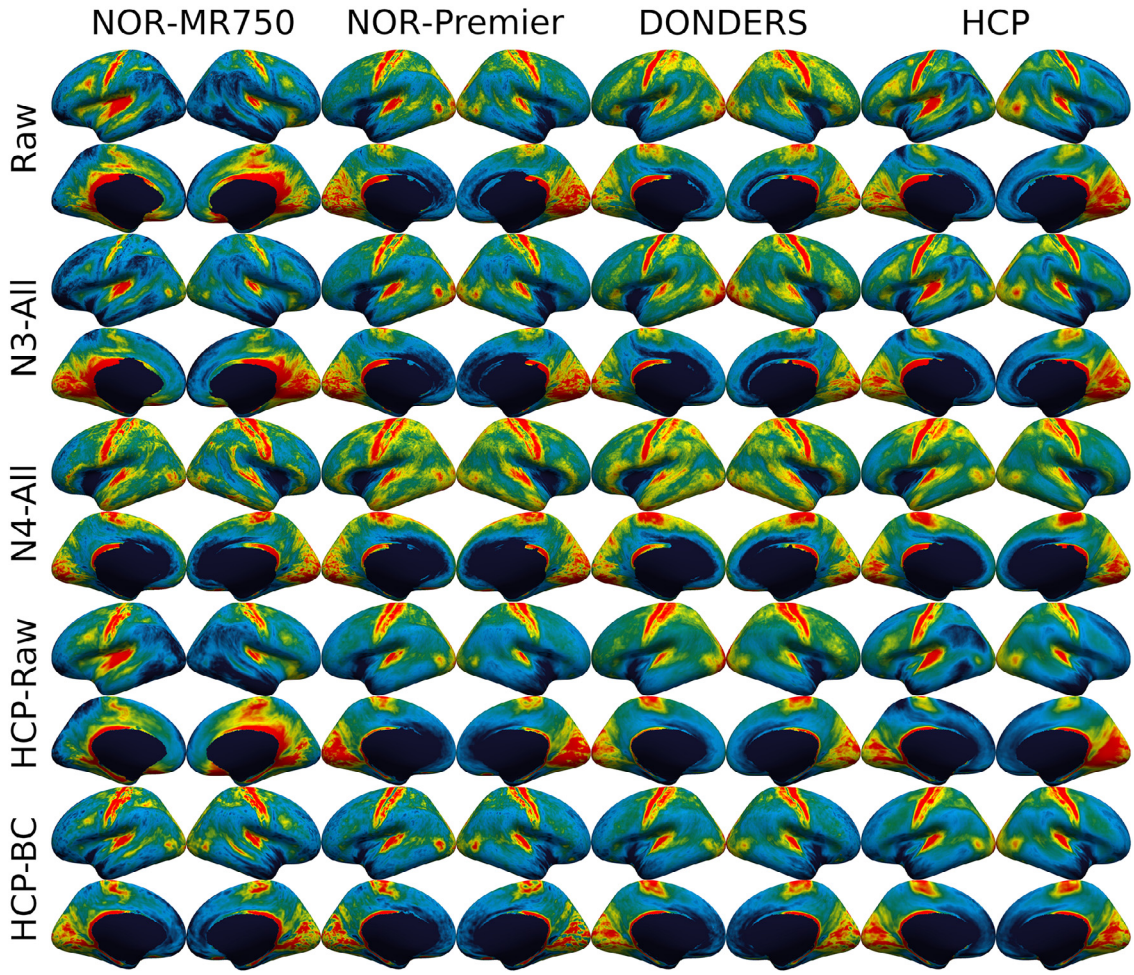


Fig. 2. Mean T1w/T2w-ratio maps in each dataset for a selection of calculation methods. Partial volume correction and surface outlier correction yielded similar maps as the raw processing method. Intensity normalisation after bias correction yielded qualitatively identical maps to those before intensity normalisation and are therefore not depicted. Colours were chosen by distributing them at five equally spaced points from the 3rd percentile to the 96th percentile, and were set to black, blue, green, yellow and red respectively.

= −0.05], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = −0.04]. There was a moderate improvement for T1w/T2w$_{N3}$ [ρ = 0.36] and T1w/T2w$_{N3\text{-}All}$ [ρ = 0.36]. The highest correlations were seen for T1w/T2w$_{N4}$ [ρ = 0.66], T1w/T2w$_{N4\text{-}All}$ [ρ =0.70], and T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.77].

Intensity normalisation did not strongly affect rank correlations with the YA-BC and the YA-B1+ reference datasets (see Tables 3-5).

3.2.2. NOR-Premier

In the NOR-Premier dataset, we found high correlations with the YA-BC reference for T1w/T2w$_{Raw}$ [ρ = 0.89], T1w/T2w$_{OC}$ [ρ = 0.89], T1w/T2w$_{PVC}$ [ρ = 0.89], T1w/T2w$_{N3}$ [ρ = 0.86], T1w/T2w$_{N3\text{-}All}$ [ρ = 0.86], T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.90], and T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.92]. Lower correlations were seen with T1w/T2w$_{N4}$ [ρ = 0.56] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.56].

We found high correlations with the YA-B1+ reference for T1w/T2w$_{Raw}$ [ρ = 0.86], T1w/T2w$_{OC}$ [ρ = 0.87], T1w/T2w$_{PVC}$ [ρ = 0.86], T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.88], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.90]. Correlations were marginally lower for T1w/T2w$_{N3}$ [ρ = 0.81] and T1w/T2w$_{N3\text{-}All}$ [ρ = 0.82]. The lowest correlations were observed for T1w/T2w$_{N4}$ [ρ = 0.73] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.73].

Intensity normalisation did not greatly affect correlations with the reference datasets, with the exception of ZS which decreased the correlations with the YA-B1+ reference to ρ = 0.77 when applied to T1w/T2w$_{Raw}$.

3.2.3. Donders

In the Donders dataset, high correlations with the YA-BC reference were seen for T1w/T2w$_{Raw}$ [ρ = 0.80], T1w/T2w$_{OC}$ [ρ = 0.80], T1w/T2w$_{PVC}$ [ρ = 0.80], T1w/T2w$_{N3}$ [ρ = 0.76], T1w/T2w$_{N3\text{-}All}$ [ρ = 0.76], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.79]. Lower correlations were seen with T1w/T2w$_{N4}$ [ρ = 0.31] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.46]. The best results were seen with T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.96].

We found high correlations with the YA-B1+ reference for T1w/T2w$_{Raw}$ [ρ = 0.96], T1w/T2w$_{OC}$ [ρ = 0.96], T1w/T2w$_{PVC}$ [ρ = 0.96], T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.91], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.97]. Correlations were lower for T1w/T2w$_{N3}$ [ρ = 0.86] and T1w/T2w$_{N3\text{-}All}$ [ρ = 0.87]. The lowest correlations were observed for T1w/T2w$_{N4}$ [ρ = 0.44] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.63].

Intensity normalisation did not greatly affect correlations with the reference datasets, with the exception of RAVEL and GMM which decreased the correlations with the YA-BC reference toρ = 0.68 and ρ = 0.18 respectively when applied to T1w/T2w$_{Raw}$.

3.2.4. HCP

In the HCP dataset, high correlations with the YA-BC reference were seen for T1w/T2w$_{Raw}$ [ρ = 0.90], T1w/T2w$_{OC}$ [ρ = 0.89], T1w/T2w$_{PVC}$ [ρ = 0.90], T1w/T2w$_{N3}$ [ρ = 0.92], T1w/T2w$_{N3\text{-}All}$ [ρ = 0.91], T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.93], and T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.97]. The lowest correlations were seen for T1w/T2w$_{N4}$ [ρ = 0.71] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.71].

Table 2
Summary statistics for processing methods without intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		Raw	PVC only	OC only	N3 only	N3-All	N4 only	N4-All	HCP-Raw	HCP-BC
Mean T1w/T2w-ratio	NOR-MR750	1.37	1.36	1.37	1.44	1.43	2.75	1.49	2.02	1.30
	NOR-Premier	1.97	1.96	1.97	2.13	2.11	1.98	2.17	3.20	1.29
	DONDERS	1.38	1.37	1.38	1.48	1.47	2.53	1.57	2.55	1.29
	HCP	1.25	1.24	1.24	1.30	1.29	1.53	1.52	1.81	1.30
	Total (Mean)	1.49	1.48	1.49	1.59	1.57	2.20	1.68	2.39	1.30
Cofficient of variation (%)	NOR-MR750	12.30	12.33	12.00	12.14	11.84	14.32	13.96	6.93	6.61
	NOR-Premier	8.82	8.84	8.45	8.69	8.33	9.25	8.88	6.23	8.48
	DONDERS	8.56	8.55	8.18	8.05	7.65	9.42	9.05	6.59	6.73
	HCP	10.41	10.45	10.11	10.17	9.88	10.70	10.43	4.88	4.32
	Total (RMS)	10.13	10.15	9.80	9.89	9.56	11.11	10.78	6.20	6.70
Intraclass correlation coefficient	NOR-MR750	0.13	0.13	0.13	0.17	0.17	0.26	0.26	0.31	0.41
	NOR-Premier	0.90	0.90	0.90	0.93	0.93	0.93	0.93	0.59	0.54
	DONDERS	0.57	0.57	0.57	0.61	0.62	0.27	0.27	0.14	0.67
	HCP	0.73	0.73	0.73	0.73	0.73	0.74	0.74	0.69	0.55
	Total (Mean)	0.58	0.58	0.58	0.61	0.61	0.55	0.55	0.43	0.54
Laterality Index (%)	NOR-MR750	8.80	8.83	8.82	−0.20	−0.13	0.90	1.01	9.59	−0.03
	NOR-Premier	1.93	1.91	1.92	0.03	0.00	0.11	0.05	1.92	−0.10
	DONDERS	0.36	0.34	0.41	0.07	0.09	−0.50	−0.43	0.65	−0.11
	HCP	−0.14	−0.13	0.02	−0.41	−0.25	−0.27	−0.11	0.14	−0.22
	Total (RMS)	4.51	4.52	4.52	0.23	0.15	0.54	0.55	4.90	0.13
Correlation with YA-BC	NOR-MR750	0.27	0.26	0.27	0.70	0.70	0.53	0.57	0.23	0.92
	NOR-Premier	0.89	0.89	0.89	0.86	0.86	0.56	0.56	0.90	0.92
	DONDERS	0.80	0.80	0.80	0.76	0.76	0.31	0.46	0.79	0.96
	HCP	0.90	0.90	0.89	0.92	0.91	0.71	0.71	0.93	0.97
	Total (Mean)	0.71	0.71	0.71	0.81	0.81	0.53	0.58	0.71	0.94
Correlation with YA-B1+	NOR-MR750	−0.05	−0.05	−0.04	0.36	0.36	0.66	0.70	−0.04	0.77
	NOR-Premier	0.86	0.86	0.87	0.81	0.82	0.73	0.73	0.90	0.88
	DONDERS	0.96	0.96	0.96	0.86	0.87	0.44	0.63	0.97	0.91
	HCP	0.72	0.72	0.72	0.79	0.80	0.80	0.82	0.77	0.88
	Total (Mean)	0.62	0.62	0.63	0.71	0.71	0.66	0.72	0.65	0.86

We found high correlations with the YA-B1+ reference for $T1w/T2w_{N3\text{-}All}$ [ρ = 0.77], $T1w/T2w_{N3}$ [ρ = 0.79], $T1w/T2w_{N3\text{-}All}$ [ρ = 0.80], $T1w/T2w_{N4}$ [ρ = 0.80], $T1w/T2w_{N4\text{-}All}$ [ρ = 0.82], and $T1w/T2w_{HCP\text{-}BC}$ [ρ = 0.88]. The lowest correlations were observed for $T1w/T2w_{Raw}$ [ρ = 0.72], $T1w/T2w_{OC}$ [ρ = 0.72], and $T1w/T2w_{PVC}$ [ρ = 0.72].

Intensity normalisation did not strongly affect rank correlations with the YA-BC reference (see Tables 3-5).

3.3. Whole-cortex coefficients of variation

See Table 2 for whole-cortex coefficients of variation, CV_{Cortex}, for each processing method without intensity normalisation grouped by dataset, and Tables 3-5 for CV_{Cortex} after intensity normalisation for $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ respectively. See Fig. 4 for box plots of subject-wise CV_{Cortex} for each processing method.

3.3.1. NOR-MR750

Total whole-cortex CVs were high in the NOR-MR750 dataset for all the non-normalised processing methods, $T1w/T2w_{Raw}$ [CV_{Cortex} = 12.30%], $T1w/T2w_{PVC}$ [CV_{Cortex} = 12.33%], $T1w/T2w_{OC}$ [CV_{Cortex} = 12.00%], $T1w/T2w_{N3}$ [CV_{Cortex} = 12.14%], and $T1w/T2w_{N3\text{-}All}$ [CV_{Cortex} = 11.84%]. Whole-cortex CVs were higher with N4ITK bias correction; $T1w/T2w_{N4}$ [CV_{Cortex} = 14.32%] and $T1w/T2w_{N4\text{-}All}$ [CV_{Cortex} = 13.96%]. The lowest CVs were seen with $T1w/T2w_{HCP\text{-}Raw}$ [CV_{Cortex} = 6.93%] and $T1w/T2w_{HCP\text{-}BC}$ [CV_{Cortex} = 6.61%].

3.3.2. NOR-Premier

For the NOR-Premier dataset, whole-cortex CVs were lower than in the NOR-MR750 and HCP datasets and similar to those of the Donders dataset, with $T1w/T2w_{Raw}$ [CV_{Cortex} = 8.82%], $T1w/T2w_{PVC}$ [CV_{Cortex} = 8.84%], $T1w/T2w_{OC}$ [CV_{Cortex} = 8.45%], $T1w/T2w_{N3}$ [CV_{Cortex} = 8.69%], $T1w/T2w_{N3\text{-}All}$ [CV_{Cortex} = 8.33%], $T1w/T2w_{N4\text{-}All}$ [CV_{Cortex} = 8.88%], and $T1w/T2w_{HCP\text{-}BC}$ [CV_{Cortex} = 8.48%]. Whole-cortex CVs were the highest with N4ITK bias correction only; $T1w/T2w_{N4}$ [CV_{Cortex} = 9.25%] and the lowest with $T1w/T2w_{HCP\text{-}Raw}$ [CV_{Cortex} = 6.23%].

3.3.3. Donders

For the Donders dataset the total whole-cortex CVs were overall lower than in the NOR-MR750 and HCP datasets, with $T1w/T2w_{Raw}$ [CV_{Cortex} = 8.56%], $T1w/T2w_{PVC}$ [CV_{Cortex} = 8.55%], $T1w/T2w_{OC}$ [CV_{Cortex} = 8.18%], $T1w/T2w_{N3}$ [CV_{Cortex} = 8.05%], and $T1w/T2w_{N3\text{-}All}$ [CV_{Cortex} = 7.65%]. Whole-cortex CVs were the highest with N4ITK bias correction; $T1w/T2w_{N4}$ [CV_{Cortex} = 9.42%] and $T1w/T2w_{N4\text{-}All}$ [CV_{Cortex} = 9.05%]. The lowest CVs were seen with $T1w/T2w_{HCP\text{-}Raw}$ [CV_{Cortex} = 6.59%] and $T1w/T2w_{HCP\text{-}BC}$ [CV_{Cortex} = 6.73%].

3.3.4. HCP

In the HCP dataset, total whole-cortex CVs were similar across most of the non-intensity normalised processing methods, with $T1w/T2w_{Raw}$ [CV_{Cortex} = 10.41%], $T1w/T2w_{PVC}$ [CV_{Cortex} = 10.45%], $T1w/T2w_{OC}$ [CV_{Cortex} = 10.11%], $T1w/T2w_{N3}$ [CV_{Cortex} = 10.17%], $T1w/T2w_{N3\text{-}All}$ [CV_{Cortex} = 9.88%], $T1w/T2w_{N4}$ [CV_{Cortex} = 10.70%], and $T1w/T2w_{N4\text{-}All}$ [CV_{Cortex} = 10.43%]. Similar to the other datasets, the lowest CVs were seen with $T1w/T2w_{HCP\text{-}Raw}$ [CV_{Cortex} = 4.88%] and $T1w/T2w_{HCP\text{-}BC}$ [CV_{Cortex} = 4.32%].

3.3.5. Intensity normalisation

For intensity normalisation with the raw processing method, the greatest improvement of CV_{Cortex} across datasets compared to

Table 3
Summary statistics for T1w/T2w$_{Raw}$ after intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for the median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		Raw - FCM	Raw - GMM	Raw - KDE	Raw - LSQ	Raw - N&U	Raw - RAVEL	Raw - WS	Raw - ZS
Mean T1w/T2w-ratio	NOR-MR750	0.53	0.58	0.57	0.94	0.69	0.84	0.84	0.97
	NOR-Premier	0.57	0.62	0.59	0.91	2.04	0.84	0.84	0.97
	DONDERS	0.59	0.60	0.69	0.79	1.41	0.87	0.86	0.97
	HCP	0.49	0.52	0.52	0.83	1.27	0.80	0.80	0.96
	Total (Mean)	0.55	0.58	0.59	0.87	1.35	0.84	0.83	0.97
Cofficient of variation (%)	NOR-MR750	8.32	8.29	8.48	8.35	17.83	5.94	6.04	5.61
	NOR-Premier	7.90	7.88	7.96	8.02	9.89	5.96	5.91	5.59
	DONDERS	7.47	7.80	10.64	7.43	13.21	5.40	5.50	4.52
	HCP	7.28	7.21	10.25	7.36	14.65	5.37	6.18	4.96
	Total (RMS)	7.75	7.81	9.40	7.80	14.18	5.68	5.91	5.19
Intraclass correlation coefficient	NOR-MR750	0.94	0.96	0.87	0.99	0.78	0.51	0.66	0.98
	NOR-Premier	0.91	0.93	0.89	0.99	0.97	0.76	0.69	0.97
	DONDERS	0.77	0.64	0.78	0.96	0.51	0.82	0.83	0.93
	HCP	0.86	0.88	0.53	0.96	0.85	0.51	0.40	0.93
	Total (Mean)	0.87	0.86	0.77	0.97	0.78	0.65	0.65	0.96
Laterality Index (%)	NOR-MR750	8.80	8.80	8.80	8.80	8.78	2.28	2.28	1.27
	NOR-Premier	1.93	1.93	1.93	1.93	1.92	0.64	0.64	0.36
	DONDERS	0.36	0.36	0.36	0.36	0.29	0.23	0.32	0.32
	HCP	−0.14	−0.07	−0.14	−0.14	−0.15	−0.01	0.01	−0.01
	Total (RMS)	4.51	4.51	4.51	4.51	4.49	1.19	1.20	0.68
Correlation with YA-BC	NOR-MR750	0.28	0.28	0.27	0.28	0.25	0.27	0.29	0.27
	NOR-Premier	0.88	0.88	0.88	0.89	0.89	0.86	0.90	0.87
	DONDERS	0.80	0.18	0.80	0.80	0.77	0.68	0.83	0.88
	HCP	0.90	0.90	0.90	0.90	0.89	0.88	0.89	0.88
	Total (Mean)	0.71	0.56	0.71	0.72	0.70	0.67	0.73	0.73
Correlation with YA-B1+	NOR-MR750	−0.04	−0.04	−0.04	−0.03	−0.06	−0.04	−0.02	−0.05
	NOR-Premier	0.88	0.88	0.88	0.87	0.87	0.88	0.88	0.77
	DONDERS	0.96	0.96	0.96	0.96	0.95	0.87	0.96	0.92
	HCP	0.72	0.72	0.72	0.72	0.71	0.71	0.73	0.72
	Total (Mean)	0.63	0.63	0.63	0.63	0.62	0.61	0.64	0.59

T1w/T2w$_{Raw}$ [CV$_{Cortex}$ = 10.13%] was seen after intensity normalisation with WhiteStripe [CV$_{Cortex}$ = 5.91%], RAVEL [CV$_{Cortex}$ = 5.68%], and Z-Score [CV$_{Cortex}$ = 5.19%]. For the other five intensity normalisation methods, some resulted in improvements while others made CVs worse with CV$_{Cortex}$ ranging from 7.80% to 14.18% after intensity normalisation.

The best results for the N3-All processing method after intensity normalisation as compared with T1w/T2w$_{N3\text{-}All}$ [CV$_{Cortex}$ = 9.56%] were seen with WhiteStripe [CV$_{Cortex}$ = 5.42%], RAVEL [CV$_{Cortex}$ = 5.46%], and Z-Score [CV$_{Cortex}$ = 5.17%]. Similarly, the best results for the N4-All processing method after intensity normalisation as compared with T1w/T2w$_{N4\text{-}All}$ [CV$_{Cortex}$ = 10.78%] were seen with WhiteStripe [CV$_{Cortex}$ = 5.45%], RAVEL [CV$_{Cortex}$ = 5.57%], and Z-Score [CV$_{Cortex}$ = 5.17%].

3.4. Regional coefficients of variation maps

See Fig. 3 for CV$_{ROI}$ for a selection of processing methods. Regional maps of CV$_{ROI}$ for the HCP-MMP atlas, showed that the regions with the greatest CV$_{ROI}$ tended to be located in the motor cortex across all datasets and in the occipital lobe for the NOR-Premier, Donders, and the HCP dataset. N3 bias correction lowered CV$_{ROI}$ in the occipital lobes in the NOR-Premier and the HCP datasets. Occipital lobe CV$_{ROI}$ remained high for T1w/T2w$_{HCP\text{-}BC}$ especially for the NOR-Premier and HCP datasets. For the Donders dataset, some frontal regions also showed relatively high CV$_{ROI}$ particularly with T1w/T2w$_{HCP\text{-}Raw}$. Some regions in the rostral anterior cingulate showed high CV$_{ROI}$ across multiple datasets and processing methods. This was likely caused by unstable parcellation performance in this region due to the proximity to the medial wall. See Supplementary Figure 6–9 for CV$_{ROI}$ maps for T1w/T2w$_{N3\text{-}All}$ and T1w/T2w$_{N4\text{-}All}$ after intensity normalisation.

3.5. Intraclass correlation coefficients

Intraclass correlation coefficients for whole-cortex median T1w/T2w-ratio values were in general low for the NOR-MR750 dataset without intensity normalisation, ranging from 0.13 for T1w/T2w$_{Raw}$ to 0.41 for T1w/T2w$_{HCP\text{-}BC}$. For the NOR-Premier dataset, ICC$_{Median}$ was high for all pipelines with a range of 0.90 to 0.93 except for T1w/T2w$_{HCP\text{-}Raw}$ and T1w/T2w$_{HCP\text{-}BC}$ with an ICC$_{Median}$ of 0.59 and 0.54 respectively. For the Donders dataset, ICC$_{Median}$ ranged from 0.67 for T1w/T2w$_{HCP\text{-}BC}$ to 0.14 for T1w/T2w$_{HCP\text{-}Raw}$. For the HCP dataset, ICC$_{Median}$ ranged from 0.55 for T1w/T2w$_{HCP\text{-}BC}$ to 0.74 for T1w/T2w$_{N4}$ and T1w/T2w$_{N4\text{-}All}$.

After intensity normalisation with both the Raw, N3-All, and N4-All processing methods, ICC$_{Median}$ improved the most with Least-Squares and Z-Score normalisation. For N3-All with Nyúl & Udupa normalisation, ICC$_{Median}$ was also high for NOR-Premier [ICC$_{Median}$ = 0.98], NOR-MR750 [ICC$_{Median}$ = 0.86] and HCP [ICC$_{Median}$ = 0.89], but lower for Donders [ICC$_{Median}$ = 0.49]. For N4-All with Nyúl & Udupa normalisation, ICC$_{Median}$ was high for all four datasets. See Tables 2-5 for median ICCs for each dataset and processing method and Fig. 5 for box plots of ICCs of mean T1w/T2w-ratio values in each ROI of the HCP-MMP atlas. See Supplementary Figures 10–14 for regional ICCs of mean T1w/T2w-ratio values in each ROI of the HCP-MMP atlas.

3.6. Laterality indices

Laterality indices showed large left-right differences for the raw processing method for the NOR-MR750 dataset with T1w/T2w$_{Raw}$ [LI = 8.80%] compared to the NOR-Premier dataset with T1w/T2w$_{Raw}$ [LI = 1.93%], the Donders dataset with T1w/T2w$_{Raw}$ [LI = 0.36%], and the HCP dataset with T1w/T2w$_{Raw}$ [LI = −0.14%], indicating

Table 4
Summary statistics for $T1w/T2w_{N3\text{-}All}$ after intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		N3-All - FCM	N3-All - GMM	N3-All - KDE	N3-All - LSQ	N3-All - N&U	N3-All - RAVEL	N3-All - WS	N3-All - ZS
Mean T1w/T2w-ratio	NOR-MR750	0.54	0.59	0.57	0.94	1.24	0.84	0.85	0.97
	NOR-Premier	0.57	0.62	0.60	0.88	1.96	0.85	0.85	0.97
	DONDERS	0.58	0.60	0.71	0.80	1.43	0.88	0.87	0.96
	HCP	0.49	0.52	0.52	0.81	1.26	0.81	0.81	0.96
	Total (Mean)	0.54	0.58	0.60	0.86	1.47	0.85	0.84	0.97
Cofficient of variation (%)	NOR-MR750	7.61	7.60	7.68	7.66	15.42	5.83	5.78	5.59
	NOR-Premier	7.30	7.30	7.40	7.54	8.71	5.91	5.74	5.58
	DONDERS	6.53	6.74	9.63	6.74	13.66	4.92	4.91	4.50
	HCP	6.53	6.46	6.52	6.55	12.22	5.10	5.19	4.95
	Total (RMS)	7.01	7.04	7.89	7.14	12.74	5.46	5.42	5.17
Intraclass correlation coefficient	NOR-MR750	0.95	0.97	0.93	0.99	0.86	0.65	0.78	0.99
	NOR-Premier	0.94	0.97	0.92	0.99	0.98	0.85	0.88	0.99
	DONDERS	0.75	0.70	0.79	0.95	0.49	0.84	0.84	0.96
	HCP	0.88	0.90	0.99	0.98	0.89	0.94	0.93	0.96
	Total (Mean)	0.88	0.89	0.91	0.98	0.80	0.82	0.86	0.97
Laterality Index (%)	NOR-MR750	−0.13	−0.13	−0.13	−0.13	−0.10	−0.17	−0.15	−0.02
	NOR-Premier	0.00	0.00	0.00	0.00	0.02	−0.06	0.00	0.01
	DONDERS	0.09	0.09	0.09	0.09	0.11	0.05	−0.05	−0.04
	HCP	−0.25	−0.24	−0.25	−0.25	−0.26	−0.10	−0.08	−0.05
	Total (RMS)	0.15	0.14	0.15	0.15	0.15	0.11	0.09	0.03
Correlation with YA-BC	NOR-MR750	0.70	0.70	0.70	0.71	0.69	0.71	0.70	0.69
	NOR-Premier	0.85	0.85	0.85	0.86	0.85	0.74	0.84	0.85
	DONDERS	0.76	0.76	0.76	0.76	0.75	0.60	0.78	0.77
	HCP	0.91	0.91	0.91	0.91	0.91	0.90	0.91	0.91
	Total (Mean)	0.81	0.81	0.81	0.81	0.80	0.74	0.81	0.80
Correlation with YA-B1+	NOR-MR750	0.37	0.37	0.37	0.37	0.35	0.38	0.36	0.35
	NOR-Premier	0.82	0.82	0.82	0.82	0.81	0.74	0.82	0.76
	DONDERS	0.87	0.87	0.87	0.87	0.86	0.72	0.85	0.85
	HCP	0.80	0.80	0.80	0.80	0.80	0.78	0.81	0.80
	Total (Mean)	0.71	0.71	0.71	0.72	0.71	0.65	0.71	0.69

large interhemispheric intensity bias in the NOR-MR750 dataset. Similarly, large laterality indices were seen in the NOR-MR750 dataset for $T1w/T2w_{PVC}$, $T1w/T2w_{OC}$, and $T1w/T2w_{HCP\text{-}Raw}$. Bias field correction had marginal effects on the laterality indices of the NOR-Premier, Donders, and HCP datasets, but reduced those of the NOR-MR750 dataset with $T1w/T2w_{N3}$ [LI = −0.20%], $T1w/T2w_{N3\text{-}All}$ [LI = −0.13%], $T1w/T2w_{N4}$ [LI = 0.90%], and $T1w/T2w_{N4\text{-}All}$ [LI = 1.01%]. Similarly, $T1w/T2w_{HCP\text{-}BC}$ decreased the laterality index to −0.03% for the NOR-MR750 dataset.

After intensity normalisation in the NOR-MR750 dataset using the Raw processing method, laterality indices decrased with RAVEL [LI = 2.28%], WhiteStripe [LI = 2.28%] and Z-Score [LI = 1.27%]. Otherwise laterality indices were only marginally affected by intensity normalisation using the Raw and N3-All processing methods. See Fig. 6 for box plots of per-subject laterality indices for processing method grouped by dataset.

3.7. Follow-up analyses

3.7.1. Analyses on high-resolution data

Spearman rank correlations with the reference datasets, ICC_{Median}, and laterality indices were similar to those observed in the main analyses on downsampled data for both the NOR-Premier and HCP datasets. Whole-cortex coefficients of variation, CV_{Cortex} were lower across datasets for $T1w/T2w_{Raw\text{-}Hires}$ [CV_{Cortex} = 7.02%] and $T1w/T2w_{HCP\text{-}Raw\text{-}Hires}$ [CV_{Cortex} = 4.71%], but higher for T1w/T2wHCP-BC-Hires [CV_{Cortex} = 9.98%]. The high CV_{Cortex} for T1w/T2wHCP-BC-Hires was driven by the NOR-Premier dataset which had a CV_{Cortex} of 13.60%. See Supplementary Table 1 for details.

3.7.2. Extraction directly from FS_LR space

Extracting ROI-wise T1w/T2w-ratio data directly from *FS_LR* space for $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ had only marginal effects on the Spearman rank correlations with the reference datasets, with coefficients differing by less than 0.01 for all datasets. See Supplementary Table 2 for the results for each processing method and dataset.

4. Discussion

Two key findings emerged from our investigation. First, the performance of the T1w/T2w-ratio in faithfully reproducing myeloarchitectonic maps is highly variable across sites and processing pipelines but can in some cases be improved with bias correction. For example, we found that for the NOR-MR750 dataset, for which reproducibility was initially low, the correlation with the YA-BC reference improved from 0.27 for $T1w/T2w_{Raw}$ to 0.70 after data-driven N3 bias correction and to 0.92 with the template-adjusted HCP-BC pipeline. Similarly, in this dataset we found an increase in the correlation with the B1+ corrected YA-B1+ dataset from −0.05 to 0.66 after data-driven N4ITK bias correction and to 0.77 with the template-adjusted HCP-BC pipeline. The second main finding was that although the test-retest reliability of the raw T1w/T2w-ratio is poor, large improvements to reliability were achieved with the use of some intensity normalisation methods, whereas other methods resulted in lower test-retest reliability.

The poor reproducibility of regional distributions seen for some datasets and processing pipelines can be an obstacle to for the myeloarchitectonic parcellation of the cerebral cortex if not sufficiently accounted for, for example using N4ITK bias correction or the template-based correction method in the HCP-MPP pipeline. This is likely caused by B1+ field inhomogeneities that are not cancelled when taking the ra-

Table 5
Summary statistics for T1w/T2w$_{N4\text{-}All}$ after intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		N4-All -FCM	N4-All -GMM	N4-All -KDE	N4-All - LSQ	N4-All - N&U	N4-All - RAVEL	N4-All - WS	N4-All - ZS
Mean T1w/T2w-ratio	NOR-MR750	0.60	0.66	0.64	1.09	3.08	0.85	0.85	0.97
	NOR-Premier	0.60	0.67	0.65	0.91	2.26	0.84	0.85	0.97
	DONDERS	0.60	0.62	0.77	0.82	1.67	0.88	0.87	0.96
	HCP	0.52	0.54	0.57	0.86	1.60	0.81	0.81	0.96
	Total (Mean)	0.58	0.62	0.65	0.92	2.15	0.85	0.85	0.97
Cofficient of variation (%)	NOR-MR750	7.59	7.60	7.72	7.77	15.73	6.20	5.85	5.59
	NOR-Premier	7.23	7.32	7.55	7.95	9.11	5.92	5.79	5.58
	DONDERS	6.26	6.52	7.09	6.59	10.88	4.90	4.75	4.50
	HCP	6.40	6.49	8.54	6.71	10.80	5.17	5.35	4.95
	Total (RMS)	6.89	7.00	7.74	7.28	11.89	5.57	5.45	5.17
Intraclass correlation coefficient	NOR-MR750	0.91	0.88	0.69	0.98	0.94	0.71	0.81	0.99
	NOR-Premier	0.94	0.92	0.78	0.98	0.97	0.76	0.86	0.99
	DONDERS	0.89	0.80	0.87	0.95	0.93	0.83	0.90	0.97
	HCP	0.89	0.86	0.85	0.96	0.96	0.83	0.86	0.96
	Total (Mean)	0.91	0.86	0.80	0.97	0.95	0.78	0.86	0.98
Laterality Index (%)	NOR-MR750	1.03	1.03	1.03	1.03	1.35	0.15	0.23	0.17
	NOR-Premier	0.07	0.07	0.07	0.07	0.09	0.00	0.04	0.03
	DONDERS	−0.38	−0.38	−0.38	−0.38	−0.81	−0.15	−0.11	−0.09
	HCP	−0.12	−0.11	−0.12	−0.12	−0.16	−0.12	−0.05	−0.04
	Total (RMS)	0.55	0.55	0.55	0.55	0.79	0.12	0.13	0.10
Correlation with YA-BC	NOR-MR750	0.58	0.58	0.58	0.58	0.49	0.46	0.56	0.61
	NOR-Premier	0.57	0.57	0.57	0.57	0.56	0.54	0.58	0.62
	DONDERS	0.47	0.47	0.47	0.47	0.46	0.35	0.51	0.49
	HCP	0.71	0.71	0.71	0.71	0.69	0.72	0.72	0.72
	Total (Mean)	0.58	0.58	0.58	0.58	0.55	0.52	0.59	0.61
Correlation with YA-B1+	NOR-MR750	0.70	0.70	0.70	0.71	0.65	0.60	0.69	0.69
	NOR-Premier	0.73	0.73	0.73	0.74	0.72	0.70	0.74	0.74
	DONDERS	0.63	0.63	0.63	0.63	0.60	0.50	0.66	0.65
	HCP	0.82	0.82	0.82	0.82	0.78	0.77	0.82	0.82
	Total (Mean)	0.72	0.72	0.72	0.72	0.69	0.64	0.73	0.72

tio. In these cases, N4ITK bias correction improves the correspondence with the B1+ adjusted reference dataset. Based on this, we propose that studies on the T1w/T2w-ratio should as a first step calculate correlations between the data and well-validated cortical myelin maps from the literature. We also recommend the use of on-scanner B1- field correction and dedicated sequences to calculate B1+ field maps in order to account for both receive and transmit field inhomogeneities. In the event that field maps are not available, our findings suggest that residual field inhomogeneities can be attenuated with data-driven bias correction where N4ITK bias correction is associated with the highest correspondence to B1+ adjusted reference data. However, care must be taken since bias correction can also reduce correspondence with the expected regional distributions depending on the intensity inhomogeneity profile of the dataset.

Low test-retest reliability is an obstacle for studies investigating group differences in the T1w/T2w-ratio or its associations with variables of interest. Some intensity normalisation methods improved test-retest reliability considerably, but others had a marginal effect and some even had the paradoxical effect of lowering reliability. Our results suggest that the best performance is given by the WhiteStripe and Z-score intensity normalisation methods.

4.1. The role of data acquisition: scanner and sequence

The best results were seen with the MPRAGE pulse sequence (NOR-Premier, Donders, HCP), independently of scanner vendor. This could indicate that MPRAGE is better suited than the BRAVO pulse sequence for cortical T1w/T2w-ratio mapping. Even so, by means of bias field correction and intensity normalisation, it was possible to achieve improved reproducibility of myeloarchitectonic distributions and test-retest reliability with BRAVO (NOR-MR750). As such, researchers using a different T1-weighted sequence than MPRAGE may achieve good T1w/T2w-ratio results with the use of appropriate post-hoc corrections in the processing pipeline.

4.2. Bias field correction

Given the susceptibility of the T1w/T2w-ratio to nonlinear field inhomogeneities, particularly those associated with the B1+ field, (Glasser et al., 2021; Glasser and Van Essen, 2011), we expected bias field correction to be one of the most influential correction factors. In line with this expectation, we found a major improvement in the correspondence between the NOR-MR750 dataset and the YA-BC dataset after N3 bias correction and a similar improvement in the correlation with the B1+ corrected YA-B1+ dataset after N4ITK bias correction. As the same type of on-scanner B1- field correction was performed in the NOR-MR750 dataset as for the NOR-Premier dataset, we ascribe these improvements to the reduction of field inhomogeneities caused by the B1+ field. However, we only saw minor improvements in the HCP dataset and in the NOR-Premier and Donders datasets the correspondence worsened with bias correction. This may indicate less initial field inhomogeneity due to scanner hardware differences or more effective field inhomogeneity correction at the image reconstruction stage in these datasets.

We found a greater improvement in the correlation between the NOR-MR750 dataset and the YA-BC reference after bias correction with N3 rather than N4ITK. For correlations with the new B1+ adjusted YA-B1+ reference, however, we found a greater improvement with N4ITK bias correction in the NOR-MR750 dataset. We observed moderately increased correlations between the HCP dataset and the YA-B1+ reference after both N3 and N4ITK bias correction, but for the other datasets both N3 and N4ITK bias correction lowered correlations. Given that the

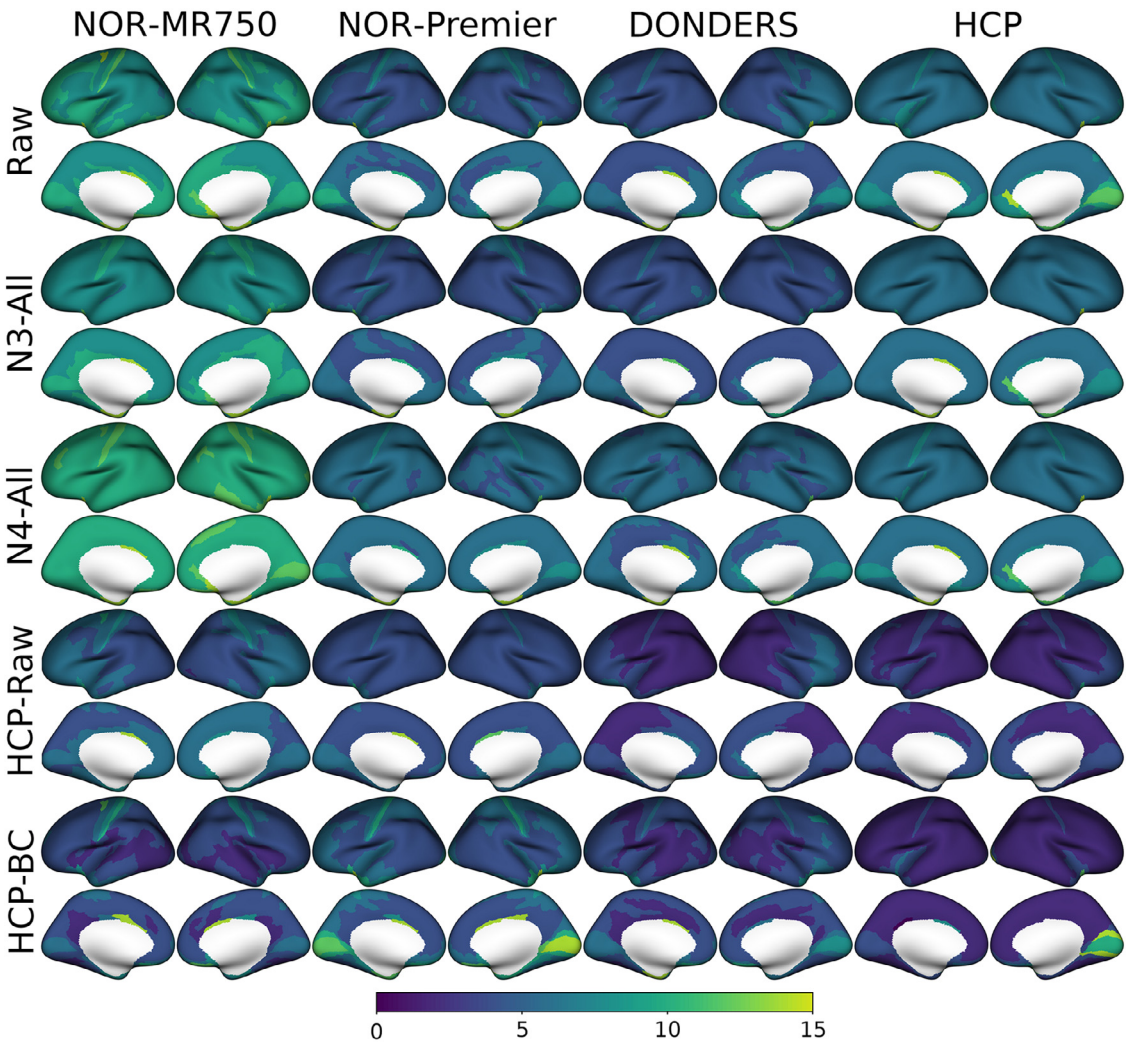


Fig. 3. CVs in each ROI of the HCP-MMP atlas for a selection of calculation methods without intensity normalisation.

YA-B1+ reference has been corrected for intensity inhomogeneities attributable to the B1+ field, we conclude that N4ITK bias correction provides the best data-driven bias correction approach of the two when residual field bias is present in the data. However, we caution against the use of either N3 or N4ITK bias correction in datasets with less intensity inhomogeneity, since bias correction may in this case worsen reproducibility. It is important that researchers seeking to use the T1w/T2w-ratio first compute correlations with well-validated datasets in order to assess the presence of intensity inhomogeneity in the data before deciding on the bias correction strategy.

The greatest reduction of the correlations with the YA-BC and YA-B1+ references in the NOR-Premier and Donders datasets was seen with N4ITK bias correction. The principal difference between the N3 and N4ITK algorithms is the B-spline smoothing strategy and the iterative optimisation approach. Whereas the N3 algorithm estimates the *total* bias field in each iteration, the N4ITK algorithm estimates the *residual* bias field from the corrected image from the previous iteration. While this has been considered to give better convergence properties for the algorithm, it may also be more aggressive than N3 bias correction. This could explain the tendency of the N4ITK algorithm to worsen correspondence with the reference datasets, especially in regions of the cortex where low T1w/T2w-ratio values were expected. It is also possible that the choice of N4ITK parameters negatively affected these results, but a thorough exploration of the N4ITK parameter space was outside the scope of the present study. Given the importance of bias field correction to the reproducibility of the NOR-MR750 dataset, and its paradoxical results of lowering correspondence with the expected T1w/T2w-ratio distributions in the NOR-Premier and Donders datasets, we encourage researchers to investigate these questions in more detail in future studies.

While $T1w/T2w_{HCP\text{-}Raw}$ neither improved the large laterality indices of the NOR-MR750 dataset nor the low correlations with the YA-BC or the YA-B1+ reference datasets, the highest correlations were seen with the $T1w/T2w_{HCP\text{-}BC}$ pipeline. The poor performance of the $T1w/T2w_{HCP\text{-}Raw}$ is likely due to the lack of bias field correction in this pipeline, with the result that residual field inhomogeneities, particularly those caused by the B1+ field, remain in the $T1w/T2w_{HCP\text{-}Raw}$ maps and consequently lead to poor correspondence with the reference datasets. In contrast, the HCP-BC pipeline employs a combined bias correction and intensity normalisation method using the smoothed difference between individual T1w/T2w-ratio maps and the population-average Conte69 template. This results in a non-linear correction of the $T1w/T2w_{HCP\text{-}Raw}$ maps where low-frequency local deviations from the Conte69 template are removed directly. Importantly, the YA-BC reference dataset was also created using the HCP-BC pipeline and it is therefore not surprising that the highest correlations to the YA-BC reference was seen with the HCP-BC pipeline. These factors complicate the comparison with the other bias correction methods, since both N3 and N4ITK bias correction are data-driven methods invoking few assumptions about the expected distributions of the final maps.

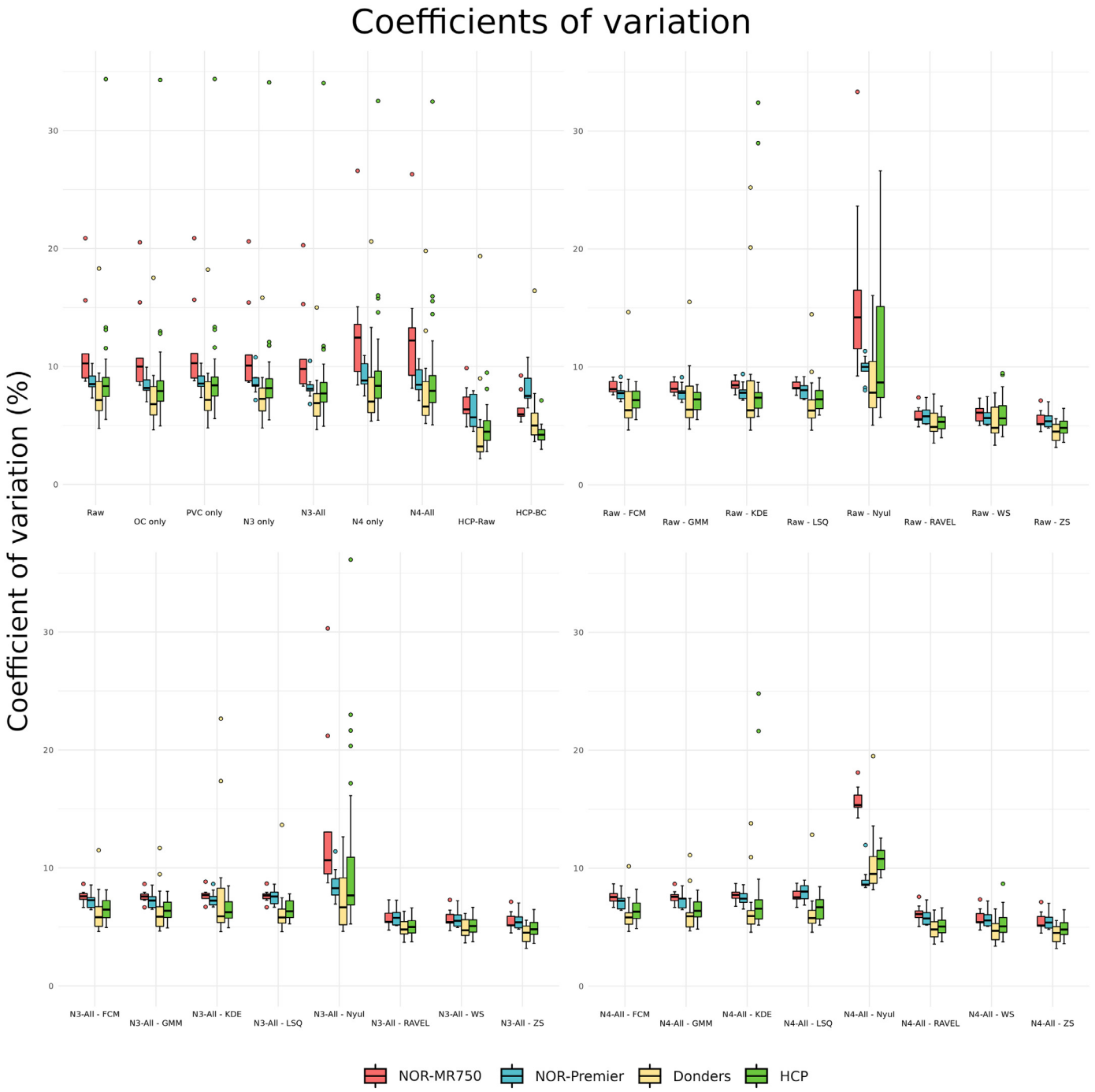


Fig. 4. Box plots of whole-cortex CVs for each participant grouped by processing method and dataset. Four outlier values (CVs > 35%) from two participants were not depicted for visualisation reasons. Their values were 38% and 46% for $T1w/T2w_{Raw}$ and 43% and 36% for $T1w/T2w_{N3\text{-}All}$ both after N&U intensity normalisation.

4.3. Test-retest reliability before intensity normalisation

We found overall low test-retest reliability of the T1w/T2w-ratio across all datasets and for most of the processing pipelines. The lowest whole-cortex CVs were achieved with $T1w/T2w_{HCP\text{-}Raw}$, while global reliability as measured with ICC_{Median} remained poor. We expected lower test-retest reliability with longer scan-rescan interval. This was in line with the lower test-retest reliability of the HCP dataset (CV_{Cortex} = 10.41% for $T1w/T2w_{Raw}$) which had a mean scan-rescan interval of 3.85 months compared to the Donders dataset which used similar acquisition parameters but acquired scan-rescan pairs on the same day and showed higher test-retest reliability (CV_{Cortex} = 8.56% for $T1w/T2w_{Raw}$). However, we also found low test-retest reliability (12.30% for $T1w/T2w_{Raw}$) with the NOR-MR750 dataset where CV_{Cortex} was calculated on the basis of same-day scan-rescan pairs in two sessions with only a two week scan-rescan interval.

Test-retest reliability did not improve after bias field correction and, even decreased after N4ITK bias correction in three of the datasets. This might indicate a drift of intensities during bias field correction, which could also be reflected in the increased means of $T1w/T2w_{N4}$ and $T1w/T2w_{N4\text{-}All}$ (Table 2) for two of the datasets. We also saw a broad range of ICC_{Median}, a measure of global test-retest agreement in the T1w/T2w-ratio maps, with values from 0.13 for the NOR-MR750 dataset to 0.90 for the NOR-Premier dataset. This variability may

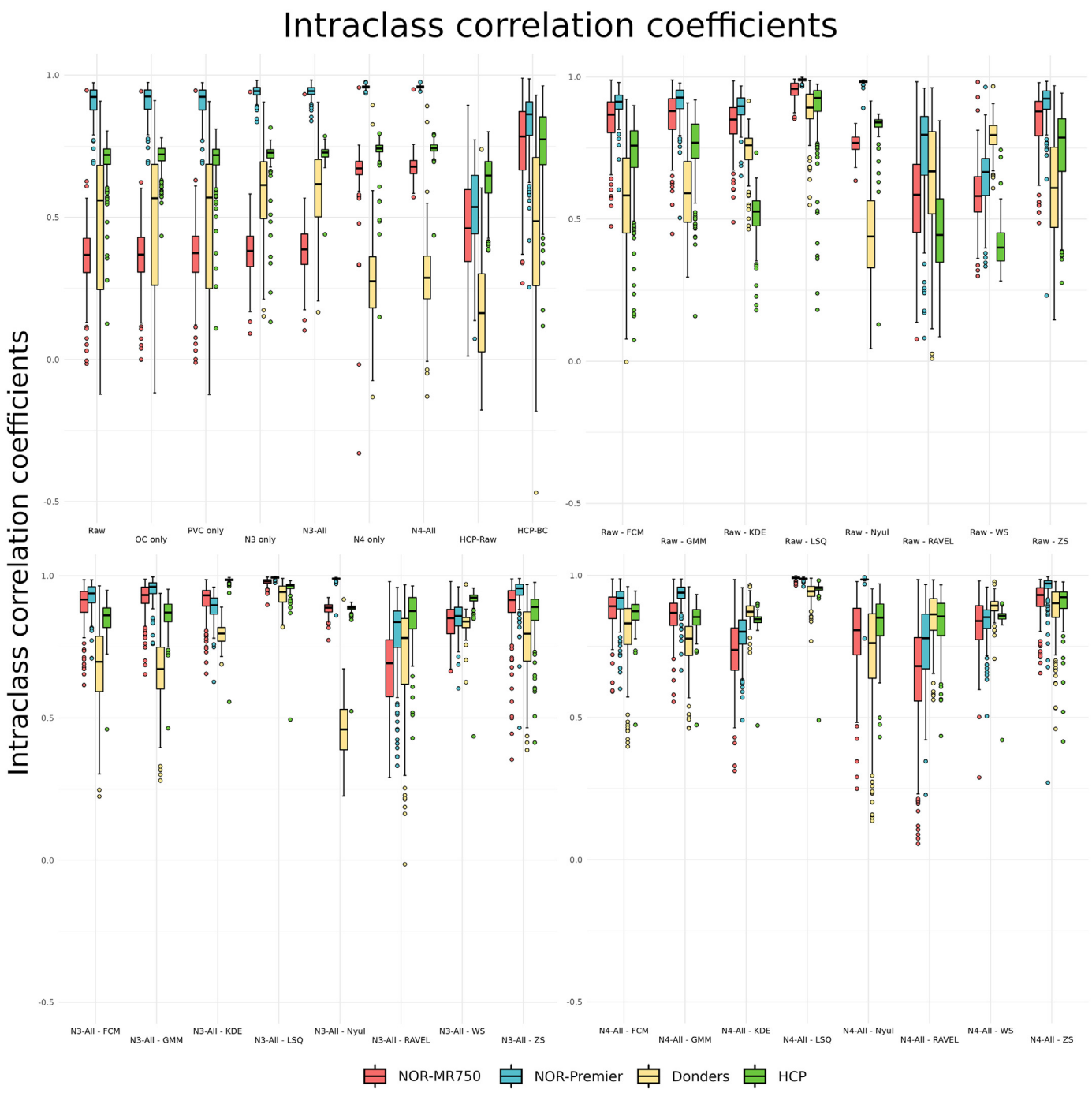


Fig. 5. Box plots of ICCs for each ROI grouped by processing method and dataset.

indicate scanner differences in individual transmit and receive gain settings determined during prescan. This may lead to global intensity drift that make *inter-subject* comparisons without intensity normalisation difficult if not impossible. It is noteworthy that some pipelines, such as $T1w/T2w_{HCP\text{-}Raw}$ decreased ICC_{Median} in three of the datasets.

4.4. Test-retest reliability after intensity normalisation

There were major differences between intensity normalisation methods. Some methods, in particular WhiteStripe and Z-score, improved the test-retest reliability considerably relative to $T1w/T2w_{Raw}$. The similarity between WhiteStripe and RAVEL is likely due to WhiteStripe being performed as part of the RAVEL procedure. Despite improvements in the total whole-cortex CVs with RAVEL, ICC_{Median} was low for the NOR-MR750 dataset at 0.71 for $T1w/T2w_{N4\text{-}All}$, whereas more consistent results were found with WhiteStripe normalisation. We observed the best numerical results with Z-Score normalisation, possibly the most straightforward method, yielding a total whole-cortex CV of 5.17% for the $T1w/T2w_{N4\text{-}All}$ pipeline. Some intensity normalisation procedures yielded poor results for test-retest reliability, in particular one previously proposed method, N&U (Nyúl and Udupa, 1999), led to total CVs across datasets of 14.18% when applied to $T1w/T2w_{Raw}$ which was higher than the CV of 10.13% for $T1w/T2w_{Raw}$.

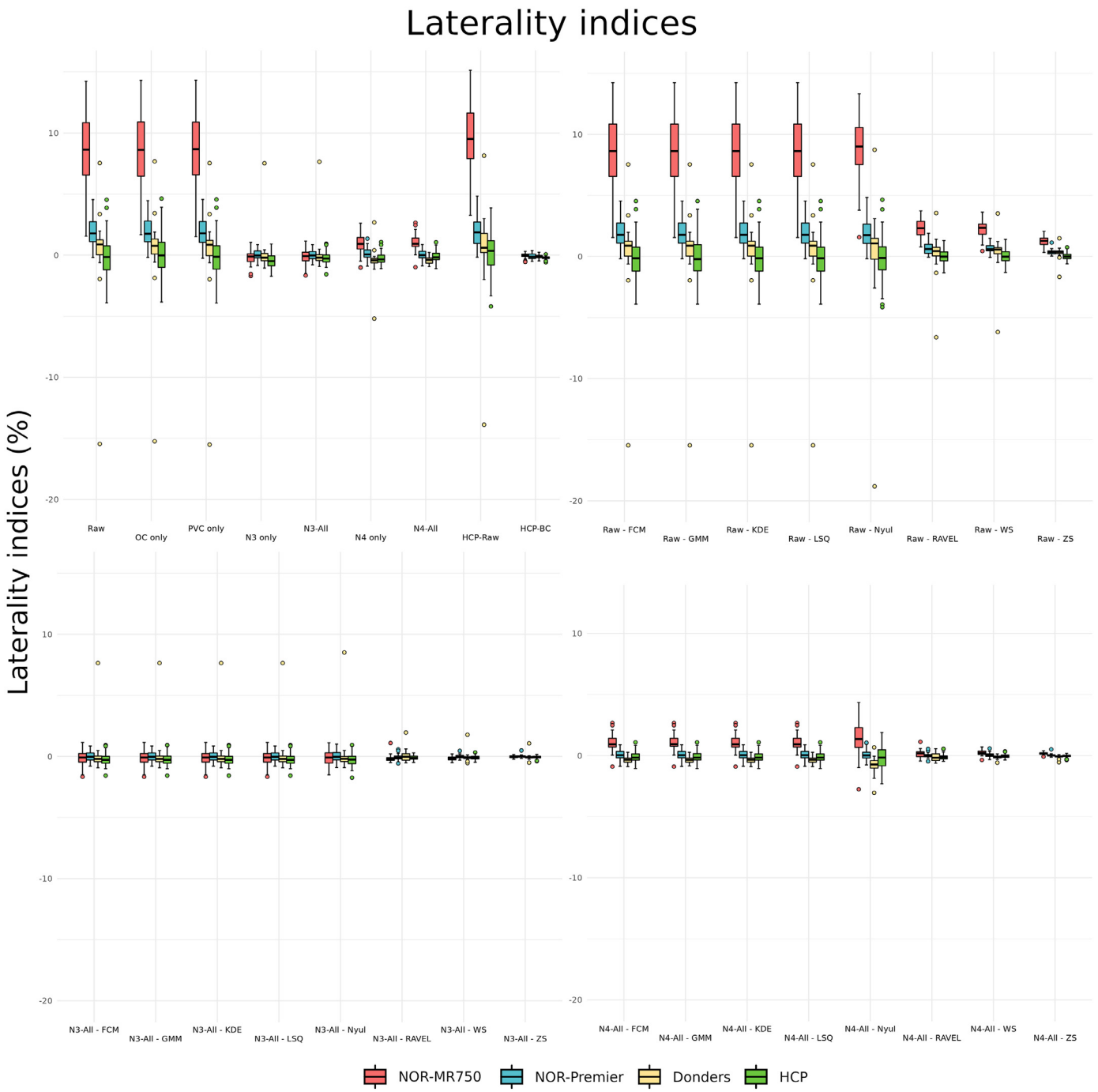


Fig. 6. Box plots of mean laterality indices for each participant grouped by processing method and dataset.

4.5. Outlier and partial volume corrections

Partial volume correction had only marginal effects on the reproducibility and reliability of T1w/T2w-ratio maps. Nevertheless, in large datasets the variable of interest may exert small but systematic partial volume effects, as in the ageing brain where the intensity contrast used to separate cortex and non-cortex is known to be affected. In such cases, partial voluming can become more influential (Shafee et al., 2015). Given these considerations, we decided to still include partial volume correction in the reliability tests. It is worth pointing out that the outlier correction method used in this study was used in the early versions of the HCP-MPP, but was later removed when it was found that sub-millimetre resolution alleviated the need for it due to fewer artefactual vertex-wise values caused by misregistration of the T1w and T2w images and surface reconstruction errors (Glasser et al., 2013). This is in line with our findings in the analyses on high-resolution data. Nonetheless, we considered outlier correction to still be relevant to evaluate in our study, given that we conformed images to 1 mm isotropic resolution to ensure generalisability of our findings to clinical studies where 1 mm resolution is still the norm. Furthermore, the unstable numerics of the T1w/T2w-ratio can yield extreme values whenever T2w image intensities are close to zero. Given that the vertex-wise method used in this study is resource intensive it may be advantageous to base outlier correction on individual T1w/T2w-ratio histograms in large datasets. Since we conformed images to 1 mm isotropic resolution, this correction may have been avoidable if high resolution images had been used.

Furthermore, since the vertex-wise method used in this study is resource intensive it may be advantageous to base outlier correction on individual T1w/T2w-ratio histograms in large datasets.

4.6. Impact of correction methods on the interpretation of results

Taken together, our findings point to the benefit of optimised T1w/T2w-ratio processing depending on the specific properties of the input data, but there are potential drawbacks, particularly for the interpretation of results of clinical analyses. Ideally, correction methods reduce both intensity bias and noise in the measurements, while retaining as much biological variation as possible. However, most intensity normalisation methods use signal variation in control regions to adjust variation in the region under study (e.g. the cerebral cortex) which introduces a dependence on intensity values in the control region. In our study, the two intensity normalisation methods with the best performance, WhiteStripe and Z-score normalisation, use variation in normal-appearing WM and the whole brain respectively, which may cause confounding by these control regions when they are used to adjust the intensities in the rest of the image.

As a practical example, consider a study of group differences in cortical T1w/T2w-ratio values in individuals who have a clinical condition where white matter myelination is also affected, for example multiple sclerosis, where normal-appearing WM has also been shown to be affected (Beer et al., 2016; Cooper et al., 2019; Granziera et al., 2021). In such cases confounding is likely and should be taken into account when the intensity normalisation method and control region is chosen. Otherwise, dependence on the control region may introduce spurious group differences in the cerebral cortex or obfuscate true effects. It is important to note here that it may not be possible to completely circumvent such limitations when employing relative non-dimensional measures such as weighted MRI intensities. This has been taken as an argument for the use of quantitative MRI techniques (Edwards et al., 2018; Weiskopf et al., 2021). Still, when such data are not available, our results suggest that intensity normalisation can improve test-retest reliability and facilitate inter-subject comparisons with conventional T1w and T2w sequences, with the caveat that researchers should examine their assumptions about the independence of their biological variables of interest with respect to control regions in order to substantiate their results.

Regarding bias field correction, it is less straightforward to predict how this affects the interpretation of results. For the N3 and N4ITK bias correction methods, low spatial frequencies are identified and removed from the image. While it is possible to constrain global intensity shifts, the dependence between local intensities within the cerebral cortex and other brain tissue is harder to assess and it is possible that individual variation in the region of interest is removed or that spurious variation is introduced with these correction methods. This issue might be addressed through the development of bias field correction methods that adjust intensities in the cerebral cortex only on the basis of the *shared* estimated field between the cerebral cortex and white matter. We encourage researchers looking to acquire data for studies on the T1w/T2w-ratio to perform on-scanner B1- field correction and to obtain dedicated B1+ field maps, for example using Actual Flip angle Imaging (Yarnykh, 2007). In the recent preprint from Glasser et al. (2021), such B1+ field maps were used to attenuate the effects of B1+ field inhomogeneity on the T1w/T2w-ratio. Interestingly, they also introduce a pseudo-transmit field correction method which relies on more commonly available spin echo and gradient echo sequences. Our results indicate that in situations where such field maps are not available, the N4ITK algorithm provides an effective alternative which can greatly improve the reproducibility of the regional distribution of the T1w/T2w-ratio values as expected from the myeloarchitectonic literature.

4.7. Impact of image resolution on results

We found that whole-cortex CVs improved when high-resolution data was used. This is in line with previous findings that higher resolution images are advantageous for T1w/T2w-ratio mapping (Glasser and Essen, 2011). This improvement is likely a result of the more accurate surface reconstruction afforded by the higher resolution data. We therefore advise that high-resolution data is used, insofar as it is possible, in studies on the T1w/T2w-ratio. However, we acknowledge that practical constraints may prevent the acquisition of high-resolution structural MRI. Importantly, longer acquisition times may increase the influence of head movement, which for clinical studies, where patients often tend to move more in the scanner than controls, can be a major confound (Reuter et al., 2015). For the mean T1w/T2w-ratios, ICCs, laterality indices, and correspondence with the reference T1w/T2w-ratio datasets, the effect of resolution was marginal. This suggests that while test-retest reliability may be improved with higher resolution data reproducibility and global T1w/T2w-ratio shifts are mainly driven by properties of the data other than resolution.

4.8. Strengths and limitations

Strengths of the present study include the use of test-retest datasets acquired on four different scanners across scanner models and vendors with a large total number of scan-rescan pairs in a test-retest context. The datasets that were included in this study are highly suited to address questions of reproducibility and reliability. We implemented a large variety of previously proposed processing methods within a standardised framework which allowed for the harmonised processing of each dataset. One such standardisation procedure was to downsample high resolution datasets in order to ensure the generalisability of our findings to typically available datasets with the more commonly used voxel resolution of 1 mm isotropic. We also included the HCP Minimal Processing Pipeline as a reference processing pipeline, which is widely used and considered state-of-the-art. Finally, each processing method was compared directly with two reference datasets in order to quantify the correspondence between T1w/T2w-ratio maps. This facilitates the generalisability of results and allows for direct comparison across processing pipelines.

As reviewed in the introduction, T1w and T2w image intensities are inherently non-quantitative as voxel intensity values do not represent direct and dimensionful measurements of biophysical properties. In this context, segmentation-based methods may be particularly powerful in providing measures that may be less sensitive to spurious inter-individual variation (Rowley et al., 2015; Viviani et al., 2017). In the present study, the focus was on the cortical T1w/T2w-ratio as an independent measure with the main focus on myeloarchitectonic parcellation and direct inter-subject comparisons. As such, questions regarding its use in segmentation-based approaches and correspondence with other measures of cortical myelin such as T1 relaxometry were not investigated. Future studies should investigate segmentation-based approaches and whether different T1w/T2w-ratio processing methods affect its validity for cortical myelin mapping through correlations with quantitative myelin measures. Notably, Shams et al. (2019) found high correlations between R1 maps and the T1w/T2w-ratio with the greatest deviations observed in regions where the B1+ field deviated the most from its nominal value. For segmentation and cortical reconstruction, we used FreeSurfer and the HCP-MPP, also based on FreeSurfer, as these are the most commonly used software suites for surface-based analyses and the cortical T1w/T2w-ratio. It is possible that other cortical segmentation methods would yield different results, but a systematic comparison of such methods was outside the scope of our study.

While the present study focused on the cortical T1w/T2w-ratio, an interesting question is how the different processing pipelines presented affect the whole-brain T1w/T2w-ratio, though the T1w/T2w-ratio in non-cortical regions may present its own set of challenges (Arshad et al., 2017; Hagiwara et al., 2018; Uddin et al., 2018, 2019). Another limitation of this study is the long scan-rescan interval of the HCP test-retest dataset (mean = 3.85 months; range = [2–5]). This was partially addressed by excluding those with scan-rescan interval greater than 6

months, but the scan-rescan interval remained greater than those of the other datasets. Finally, the datasets included in this study were composed of healthy individuals and it is possible that the outcomes might be different when applying these methods in clinical studies.

Conclusion

We recommend that future studies using the T1w/T2w-ratio for myeloarchitectonic parcellation assess the reproducibility of cortical myelin distributions by direct comparison with datasets that correspond closely with the expected myeloarchitectonic distributions, such as the YA-B1+ dataset, or atlases based on histological cortical myelin maps. For researchers planning to acquire data for studies on the T1w/T2w-ratio, we recommend that they carefully assess the scan-rescan stability of the head coil, perform on-scanner B1- field correction, and acquire scans suitable for estimating the B1+ field to correct for transmit field inhomogeneities. However, in the presence of field inhomogeneities obscuring the expected regional distributions of the T1w/T2w-ratio, bias field correction should be used and its performance quantified. Our findings suggest improved performance with the N4ITK algorithm, but this depends on the inhomogeneity profile of the specific dataset, in particular that of the B1+ field. We also found that high resolution (< 1 mm) data performed better than data downsampled to 1 mm resolution, and we recommend that images used for T1w/T2w-ratio mapping are acquired at high resolution whenever possible. We demonstrated that the test-retest reliability of the raw T1w/T2w-ratio is poor, which reduces the ability to test group differences or associations with clinical variables. Intensity normalisation methods can be used to improve reliability; the caveat being that choice of method may also affect the interpretation of results.

Data availability statement

The Donders dataset was previously used for a comparison of the T1w/T2w-ratio with MP2RAGE-based R1 mapping (Shams et al., 2019) and is available at the Donders Institute for Brain, Cognition and Behaviour repository (https://data.donders.ru.nl/). The HCP dataset is available from the ConnectomeDB (https://db.humanconnectome.org/). For the NOR-MR750 and the NOR-Premier datasets, access to the T1w and T2w images can be arranged by contacting the corresponding author and filling out a standard data sharing agreement, including a study description, given that the necessary ethics permissions have been secured.

Declaration of Competing Interest

OAA has received a speaker honorarium from Lundbeck and is a consultant for HealthLytix. Otherwise, no other declarations of interest.

Credit authorship contribution statement

Stener Nerland: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Funding acquisition. **Kjetil N. Jørgensen:** Conceptualization, Methodology, Writing – original draft. **Wibeke Nordhøy:** Investigation, Writing – original draft, Supervision. **Ivan I. Maximov:** Writing – original draft, Supervision. **Robin A.B. Bugge:** Investigation, Writing – original draft. **Lars T. Westlye:** Writing – original draft. **Ole A. Andreassen:** Writing – original draft, Resources. **Oliver M. Geier:** Writing – original draft, Supervision, Project administration. **Ingrid Agartz:** Resources, Writing – original draft, Supervision, Project administration, Funding acquisition.

Acknowledgements

This work was funded by the South-Eastern Norway Regional Health Authority; Grant No. 2017-097 and 2019-104, and the Research Council of Norway; Grant No. 223273. The Donders dataset was provided by The Donders Institute for Brain, Cognition and Behaviour (https://www.ru.nl/donders/) supported by the European FP7 program - Advanced Brain Imaging with MRI (ABRIM) and additional unprocessed data was kindly provided by Zahra Shams. The HCP dataset was provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell center for Systems Neuroscience at Washington University.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118709.

References

Arshad, M., Stanley, J.A., Raz, N., 2017. Test–retest reliability and concurrent validity of in vivo myelin content indices: myelin water fraction and calibrated T1w/T2w image ratio. Hum. Brain Mapp. 38 (4), 1780–1790. doi:10.1002/hbm.23481.

Beer, A., Biberacher, V., Schmidt, P., Righart, R., Buck, D., Berthele, A., Kirschke, J., Zimmer, C., Hemmer, B., Mühlau, M., 2016. Tissue damage within normal appearing white matter in early multiple sclerosis: assessment by the ratio of T1- and T2-weighted MR image intensity. J. Neurol. 263 (8), 1495–1502. doi:10.1007/s00415-016-8156-6.

Chen, W., Giger, M.L., 2004. A fuzzy c-means (FCM) based algorithm for intensity inhomogeneity correction and segmentation of MR images. In: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), 2, pp. 1307–1310. doi:10.1109/ISBI.2004.1398786.

Clark, V.P., Courchesne, E., Grafe, M., 1992. In vivo myeloarchitectonic analysis of human striate and extrastriate cortex using magnetic resonance imaging. Cereb. Cortex 2 (5), 417–424. doi:10.1093/cercor/2.5.417.

Cooper, G., Finke, C., Chien, C., Brandt, A.U., Asseyer, S., Ruprecht, K., Bellmann-Strobl, J., Paul, F., Scheel, M., 2019. Standardization of T1w/T2w ratio improves detection of tissue damage in multiple sclerosis. Front. Neurol. 10. doi:10.3389/fneur.2019.00334.

de Boor, C., 1972. On calculating with B-splines. J. Approx. Theory 6 (1), 50–62. doi:10.1016/0021-9045(72)90080-9.

Edwards, L.J., Kirilina, E., Mohammadi, S., Weiskopf, N., 2018. Microstructural imaging of human neocortex in vivo. Neuroimage 182, 184–206. doi:10.1016/j.neuroimage.2018.02.055.

Eickhoff, S., Walters, N.B., Schleicher, A., Kril, J., Egan, G.F., Zilles, K., Watson, J.D.G., Amunts, K., 2004. High-resolution MRI reflects myeloarchitecture and cytoarchitecture of human cerebral cortex. Hum. Brain Mapp. 24 (3), 206–215. doi:10.1002/hbm.20082.

Fortin, J.-P., Sweeney, E.M., Muschelli, J., Crainiceanu, C.M., Shinohara, R.T., 2016. Removing inter-subject technical variability in magnetic resonance imaging studies. Neuroimage 132, 198–212. doi:10.1016/j.neuroimage.2016.02.036.

Ganzetti, M., Wenderoth, N., Mantini, D., 2014. Whole brain myelin mapping using T1- and T2-weighted MR imaging data. Front. Hum. Neurosci. 8. doi:10.3389/fnhum.2014.00671.

Ghassemi, R., Brown, R., Narayanan, S., Banwell, B., Nakamura, K., Arnold, D.L., 2015. Normalization of white matter intensity on T1-weighted images of patients with acquired central nervous system demyelination. J. Neuroimaging 25 (2), 184–190. doi:10.1111/jon.12129.

Glasser, M.F., Coalson, T.S., Harms, M.P., Baum, G.L., Autio, J.A., Auerbach, E.J., Xu, J., Greve, D.N., Yacoub, E., Essen, D.C.V., Bock, N.A., & Hayashi, T. (2021). *Transmit field bias correction of T1w/T2w Myelin maps* (p. 2021.08.08.455570). https://doi.org/10.1101/2021.08.08.455570

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536 (7615), 171–178. doi:10.1038/nature18933.

Glasser, M.F., Essen, D.C.V., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J. Neurosci. 31 (32), 11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011.

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., 2013. The minimal pre-processing pipelines for the Human Connectome Project. Neuroimage 80, 105–124. doi:10.1016/j.neuroimage.2013.04.127.

Glüer, C.-C., Blake, G., Lu, Y., Blunt1, B.A., Jergas1, M., Genant1, H.K., 1995. Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. Osteoporos. Int. 5 (4), 262–270. doi:10.1007/BF01774016.

Granziera, C., Wuerfel, J., Barkhof, F., Calabrese, M., De Stefano, N., Enzinger, C., Evangelou, N., Filippi, M., Geurts, J.J.G., Reich, D.S., Rocca, M.A., Ropele, S., Rovira, À., Sati, P., Toosy, A.T., Vrenken, H., Gandini Wheeler-Kingshott, C.A.M., Kappos, L.the MAGNIMS Study Group, 2021. Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis. Brain 144 (5), 1296–1311. doi:10.1093/brain/awab029.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48 (1), 63–72. doi:10.1016/j.neuroimage.2009.06.060.

Grydeland, H., Vértes, P.E., Váša, F., Romero-Garcia, R., Whitaker, K., Alexander-Bloch, A.F., Bjørnerud, A., Patel, A.X., Sederevičius, D., Tamnes, C.K., Westlye, L.T., White, S.R., Walhovd, K.B., Fjell, A.M., Bullmore, E.T., 2019. Waves of maturation and senescence in micro-structural MRI markers of human cortical myelination over the lifespan. Cereb. Cortex 29 (3), 1369–1381. doi:10.1093/cercor/bhy330.

Grydeland, H., Walhovd, K.B., Tamnes, C.K., Westlye, L.T., Fjell, A.M., 2013. Intracortical myelin links with performance variability across the human lifespan: results from T1- and T2-weighted MRI myelin mapping and diffusion tensor imaging. J. Neurosci. 33 (47), 18618–18630. doi:10.1523/JNEUROSCI.2811-13.2013.

Hagiwara, A., Hori, M., Kamagata, K., Warntjes, M., Matsuyoshi, D., Nakazawa, M., Ueda, R., Andica, C., Koshino, S., Maekawa, T., Irie, R., Takamura, T., Kumamaru, K.K., Abe, O., Aoki, S., 2018. Myelin measurement: comparison between simultaneous tissue relaxometry, magnetization transfer saturation index, and T1w/T2w ratio methods. Sci. Rep. 8 (1), 10554. doi:10.1038/s41598-018-28852-6.

Ishida, T., Donishi, T., Iwatani, J., Yamada, S., Takahashi, S., Ukai, S., Shinosaki, K., Terada, M., Kaneoke, Y., 2017. Elucidating the aberrant brain regions in bipolar disorder using T1-weighted/T2-weighted magnetic resonance ratio images. Psychiatry Res.: Neuroimaging 263, 76–84. doi:10.1016/j.pscychresns.2017.03.006.

Iwatani, J., Ishida, T., Donishi, T., Ukai, S., Shinosaki, K., Terada, M., Kaneoke, Y., 2015. Use of T1-weighted/T2-weighted magnetic resonance ratio images to elucidate changes in the schizophrenic brain. Brain Behav. 5 (10). doi:10.1002/brb3.399.

Koenig, S.H., 1991. Cholesterol of myelin is the determinant of gray-white contrast in MRI of brain. Magn. Reson. Med. 20 (2), 285–291. doi:10.1002/mrm.1910200210.

Koenig, S.H., Brown, R.D., Spiller, M., Lundbom, N., 1990. Relaxometry of brain: why white matter appears bright in MRI. Magn. Reson. Med. 14 (3), 482–495. doi:10.1002/mrm.1910140306.

Matheson, G.J., 2019. We need to talk about reliability: making better use of test-retest studies for study design and interpretation. PeerJ 7. doi:10.7717/peerj.6918.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1 (1), 30–46. doi:10.1037/1082-989X.1.1.30.

Misaki, M., Savitz, J., Zotev, V., Phillips, R., Yuan, H., Young, K.D., Drevets, W.C., Bodurka, J., 2015. Contrast enhancement by combining T1- and T2-weighted structural brain MR images. Magn. Reson. Med. 74 (6), 1609–1620. doi:10.1002/mrm.25560.

Nakamura, K., Chen, J.T., Ontaneda, D., Fox, R.J., Trapp, B.D., 2017. T1-/T2-weighted ratio differs in demyelinated cortex in multiple sclerosis. Ann. Neurol. 82 (4), 635–639. doi:10.1002/ana.25019.

Nieuwenhuys, R., Broere, C.A.J., 2017. A map of the human neocortex showing the estimated overall myelin content of the individual architectonic areas based on the studies of Adolf Hopf. Brain Struct. Funct. 222 (1), 465–480. doi:10.1007/s00429-016-1228-7.

Norbom, L.B., Rokicki, J., Alnæs, D., Kaufmann, T., Doan, N.T., Andreassen, O.A., Westlye, L.T., Tamnes, C.K., 2020. Maturation of cortical microstructure and cognitive development in childhood and adolescence: a T1w/T2w ratio MRI study. Hum. Brain Mapp. 41 (16), 4676–4690. doi:10.1002/hbm.25149.

Nyúl, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn. Reson. Med. 42 (6), 1072–1081 10.1002/(SICI)1522-2594(199912)42:6<1072::AID–MRM11>3.0.CO;2-M.

Pelkmans, W., Dicks, E., Barkhof, F., Vrenken, H., Scheltens, P., van der Flier, W.M., Tijms, B.M., 2019. Gray matter T1-w/T2-w ratios are higher in Alzheimer's disease. Hum. Brain Mapp. 40 (13), 3900–3909. doi:10.1002/hbm.24638.

Preziosa, P., Bouman, P.M., Kiljan, S., Steenwijk, M.D., Meani, A., Pouwels, P.J., Rocca, M.A., Filippi, M., Geurts, J.J.G., Jonkman, L.E., 2021. Neurite density explains cortical T1-weighted/T2-weighted ratio in multiple sclerosis. J. Neurol., Neurosurg. Psychiatry 92 (7), 790–792. doi:10.1136/jnnp-2020-324391.

Reinhold, J.C., Dewey, B.E., Carass, A., Prince, J.L., 2019. Evaluating the impact of intensity normalization on MR image synthesis. In: Proceedings of SPIE–the International Society for Optical Engineering, p. 10949. doi:10.1117/12.2513089.

Reuter, M., Tisdall, M.D., Qureshi, A., Buckner, R.L., van der Kouwe, A.J.W., Fischl, B., 2015. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. Neuroimage 107, 107–115. doi:10.1016/j.neuroimage.2014.12.006.

Righart, R., Biberacher, V., Jonkman, L.E., Klaver, R., Schmidt, P., Buck, D., Berthele, A., Kirschke, J.S., Zimmer, C., Hemmer, B., Geurts, J.J.G., Mühlau, M., 2017. Cortical pathology in multiple sclerosis detected by the T1/T2-weighted ratio from routine magnetic resonance imaging. Ann. Neurol. 82 (4), 519–529. doi:10.1002/ana.25020.

Rokicki, J., Wolfers, T., Nordhøy, W., Tesli, N., Quintana, D.S., Alnæs, D., Richard, G., de Lange, A.G., Lund, M.J., Norbom, L., Agartz, I., Melle, I., Nærland, T., Selbæk, G., Persson, K., Nordvik, J.E., Schwarz, E., Andreassen, O.A., Kaufmann, T., Westlye, L.T., 2020. Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. Hum. Brain Mapp. 42 (6), 1714–1726. doi:10.1002/hbm.25323.

Rowley, C.D., Bazin, P.-L., Tardif, C.L., Sehmbi, M., Hashim, E., Zaharieva, N., Minuzzi, L., Frey, B.N., Bock, N.A., 2015. Assessing intracortical myelin in the living human brain using myelinated cortical thickness. Front. Neurosci. 9, 396. doi:10.3389/fnins.2015.00396.

Rowley, C.D., Tabrizi, S.J., Scahill, R.I., Leavitt, B.R., Roos, R.A.C., Durr, A., Bock, N.A., 2018. Altered intracortical T1-weighted/T2-weighted ratio signal in Huntington's disease. Front. Neurosci. 12. doi:10.3389/fnins.2018.00805.

Sereno, M.I., Lutti, A., Weiskopf, N., Dick, F., 2013. Mapping the human cortical surface by combining quantitative T1 with retinotopy. Cereb. Cortex (New York, NY) 23 (9), 2261–2268. doi:10.1093/cercor/bhs213.

Shafee, R., Buckner, R.L., Fischl, B., 2015. Gray matter myelination of 1555 human brains using partial volume corrected MRI images. Neuroimage 105, 473–485. doi:10.1016/j.neuroimage.2014.10.054.

Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T., 2011. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Med. Image Anal. 15 (2), 267–282. doi:10.1016/j.media.2010.12.003.

Shams, Z., Norris, D.G., Marques, J.P., 2019. A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. PLoS ONE 14 (7). doi:10.1371/journal.pone.0218089.

Shinohara, R.T., Sweeney, E.M., Goldsmith, J., Shiee, N., Mateen, F.J., Calabresi, P.A., Jarso, S., Pham, D.L., Reich, D.S., Crainiceanu, C.M., 2014. Statistical normalization techniques for magnetic resonance imaging. NeuroImage : Clin. 6, 9–19. doi:10.1016/j.nicl.2014.08.008.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17 (1), 87–97. doi:10.1109/42.668698.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320. doi:10.1109/TMI.2010.2046908.

Tzourio-Mazoyer, N., Maingault, S., Panzieri, J., Pepe, A., Crivello, F., Mazoyer, B., 2019. Intracortical myelination of Heschl's gyrus and the planum temporale varies with heschl's duplication pattern and rhyming performance: an investigation of 440 healthy volunteers. Cereb. Cortex 29 (5), 2072–2083. doi:10.1093/cercor/bhy088.

Uddin, M.N., Figley, T.D., Marrie, R.A., Figley, C.R., 2018. Can T1w/T2w ratio be used as a myelin-specific measure in subcortical structures? Comparisons between FSE-based T1w/T2w ratios, GRASE-based T1w/T2w ratios and multi-echo GRASE-based myelin water fractions. NMR Biomed. 31 (3), e3868. doi:10.1002/nbm.3868.

Uddin, M.N., Figley, T.D., Solar, K.G., Shatil, A.S., Figley, C.R., 2019. Comparisons between multi-component myelin water fraction, T1w/T2w ratio, and diffusion tensor imaging measures in healthy human brain structures. Sci. Rep. 9 (1), 2500. doi:10.1038/s41598-019-39199-x.

Uğurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., … Yacoub, E., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. Neuroimage 80, 80–104. doi:10.1016/j.neuroimage.2013.05.012.

Valk, S.L., Xu, T., Margulies, D.S., Masouleh, S.K., Paquola, C., Goulas, A., Kochunov, P., Smallwood, J., Yeo, B.T.T., Bernhardt, B.C., Eickhoff, S.B., 2020. Shaping brain structure: genetic and phylogenetic axes of macroscale organization of cortical thickness. Sci. Adv. 6 (39). doi:10.1126/sciadv.abb3417.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041.

Viviani, R., Stöcker, T., Stingl, J.C., 2017. Multimodal FLAIR/MPRAGE segmentation of cerebral cortex and cortical myelin. Neuroimage 152, 130–141. doi:10.1016/j.neuroimage.2017.02.054.

Waehnert, M.D., Dinse, J., Weiss, M., Streicher, M.N., Waehnert, P., Geyer, S., Turner, R., Bazin, P.-L., 2014. Anatomically motivated modeling of cortical laminae. Neuroimage 93, 210–220. doi:10.1016/j.neuroimage.2013.03.078.

Walters, N.B., Egan, G.F., Kril, J.J., Kean, M., Waley, P., Jenkinson, M., Watson, J.D.G., 2003. In vivo identification of human cortical areas using high-resolution MRI: an approach to cerebral structure–function correlation. In: Proceedings of the National Academy of Sciences of the United States of America, 100, pp. 2981–2986. doi:10.1073/pnas.0437896100.

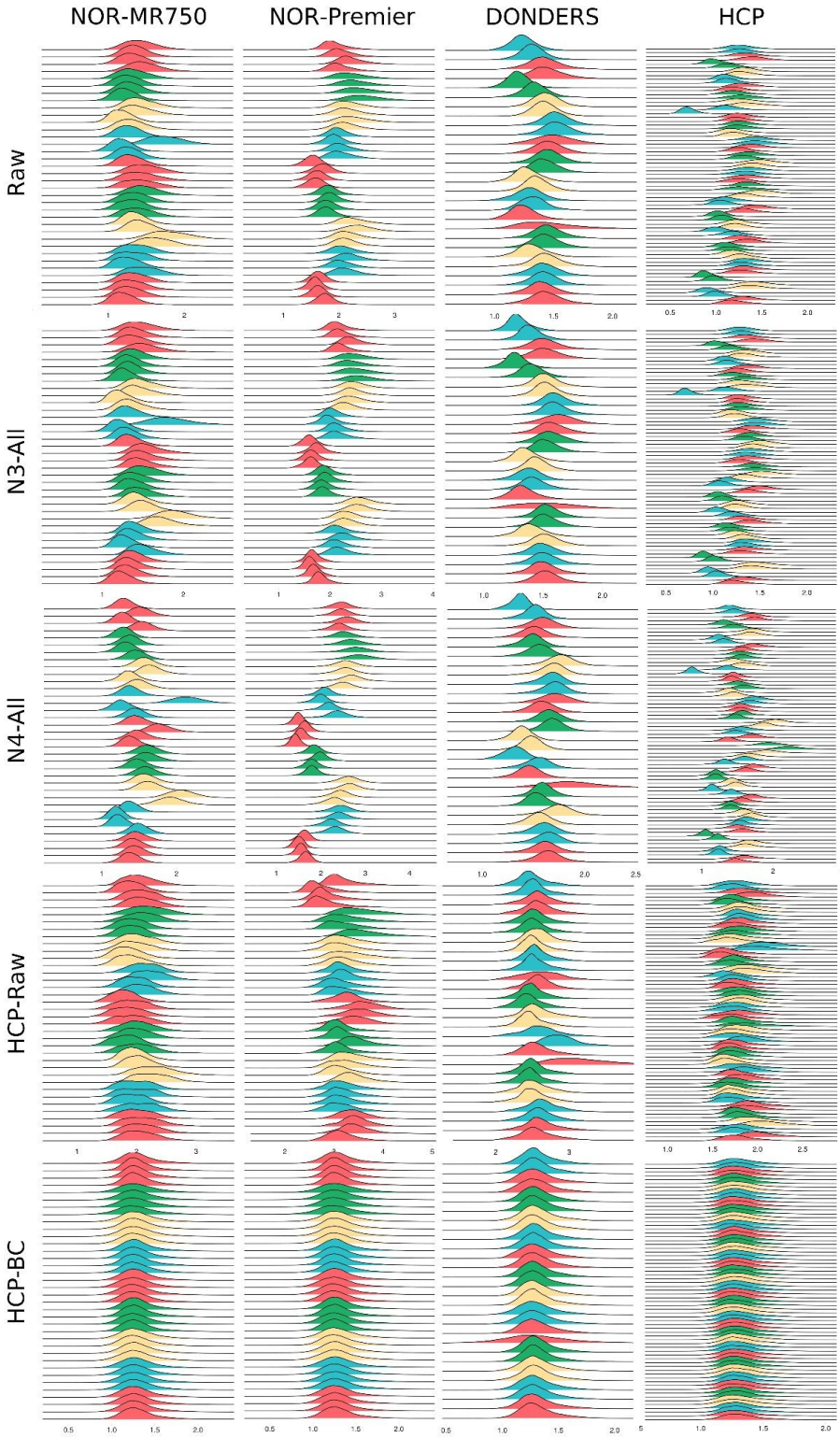
Wang, X.-J., 2020. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nat. Rev. Neurosci. 21 (3), 169–178. doi:10.1038/s41583-020-0262-x.

Weiskopf, N., Edwards, L.J., Helms, G., Mohammadi, S., Kirilina, E., 2021. Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. Nat. Rev. Phys. 3 (8), 570–588. doi:10.1038/s42254-021-00326-1.

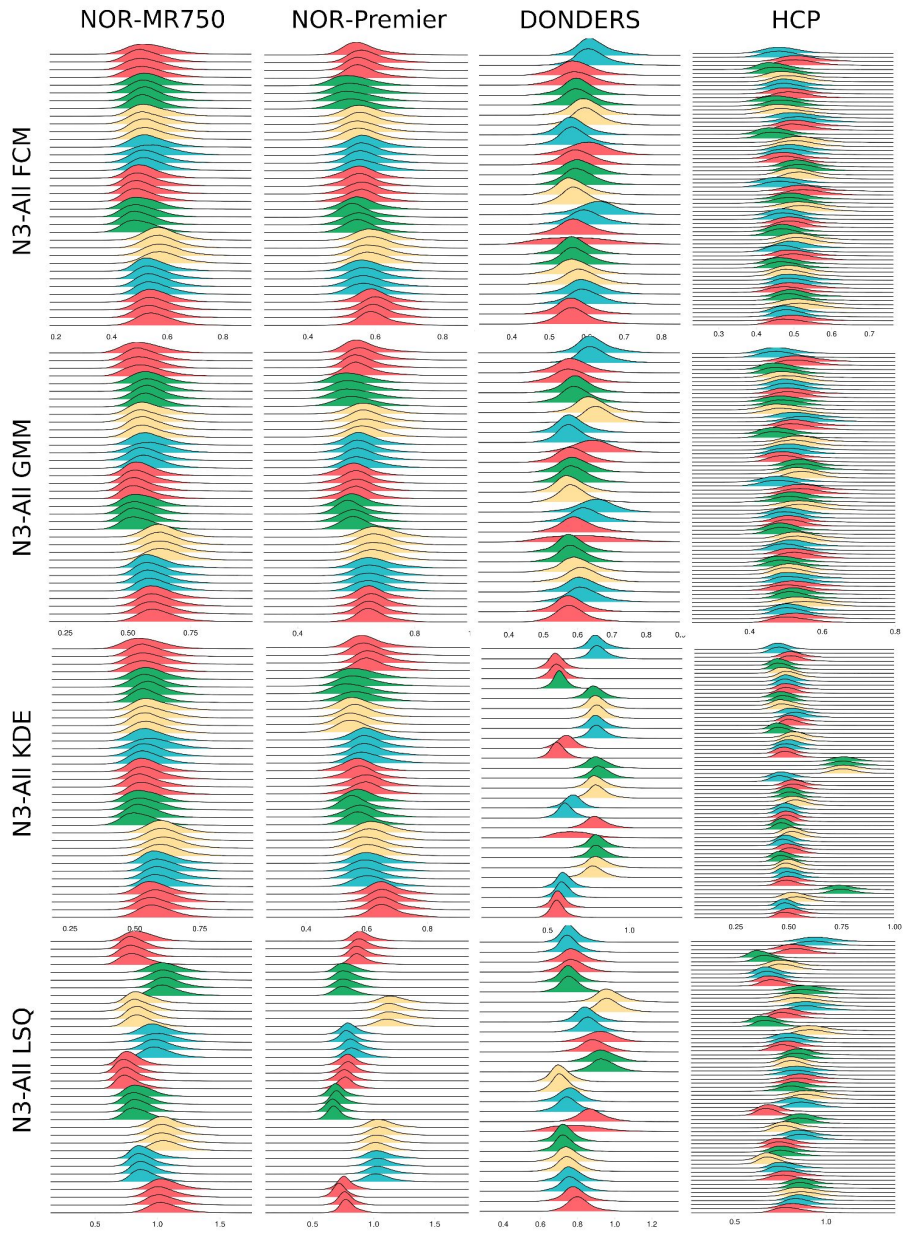
Yarnykh, V.L., 2007. Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magn. Reson. Med. 57 (1), 192–200. doi:10.1002/mrm.21120.

Wang, L. (2021). geodesic (https://www.mathworks.com/matlabcentral/fileexchange/6522-geodesic), MATLAB central file exchange. Retrieved February 13, 2021.

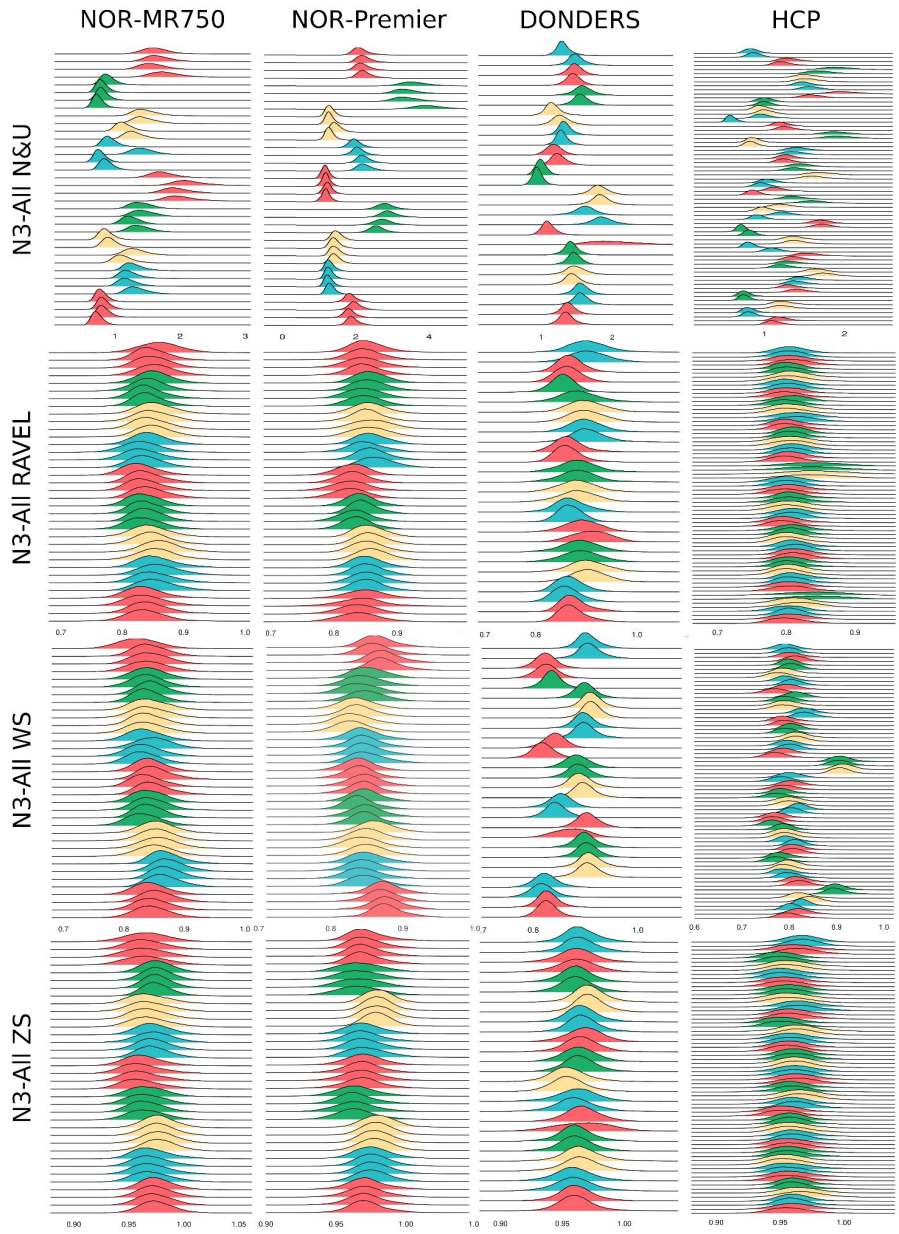
Supplementary Materials



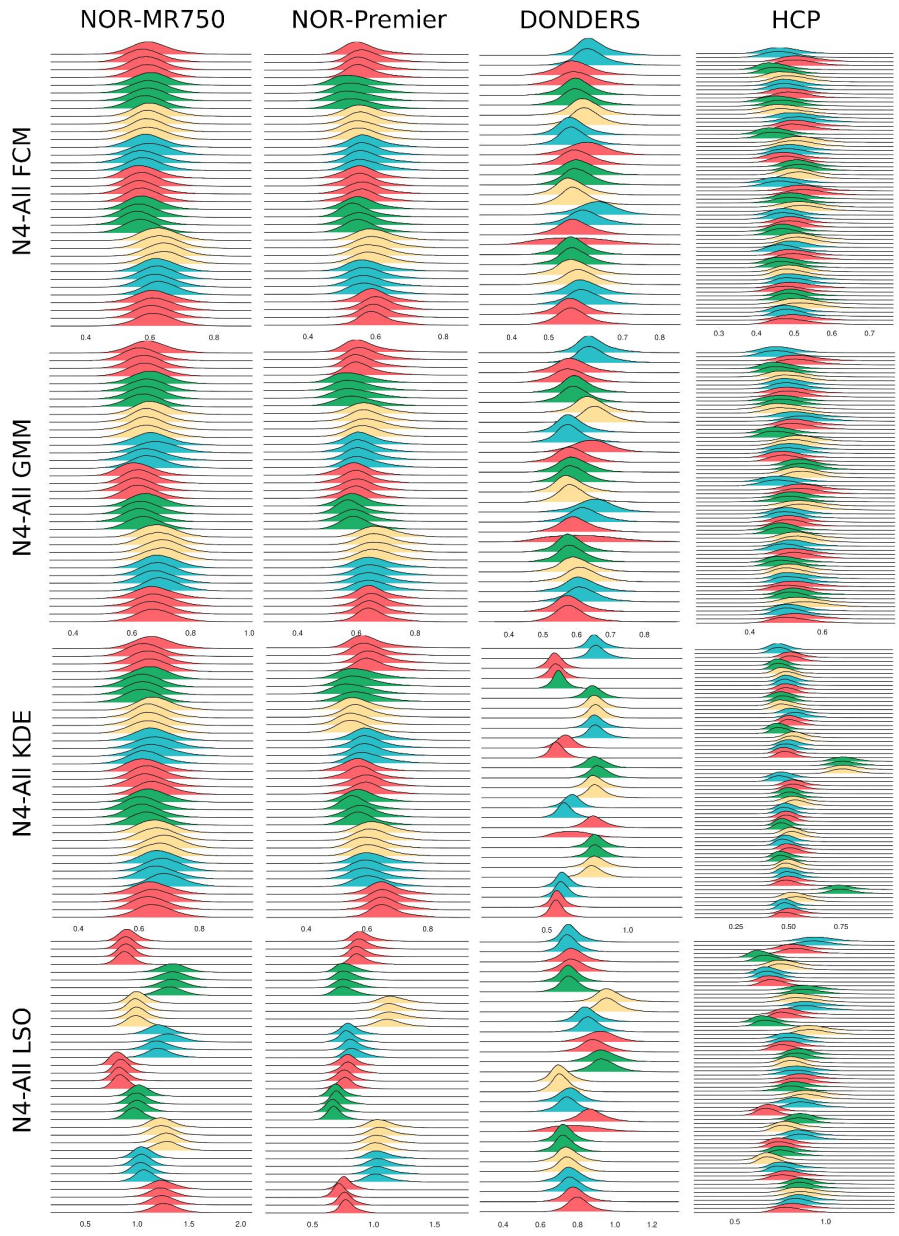
Supplementary Figure 1. T1w/T2w-ratio distributions for each participant for a selection of processing methods without intensity normalisation. Adjacent distributions with the same colour represent each acquisition for that participant. I.e. two sessions with two test-retest acquisitions for each NOR-MR750 and NOR-Premier participant, and a pair of test-retest acquisitions for each Donders and HCP participant.



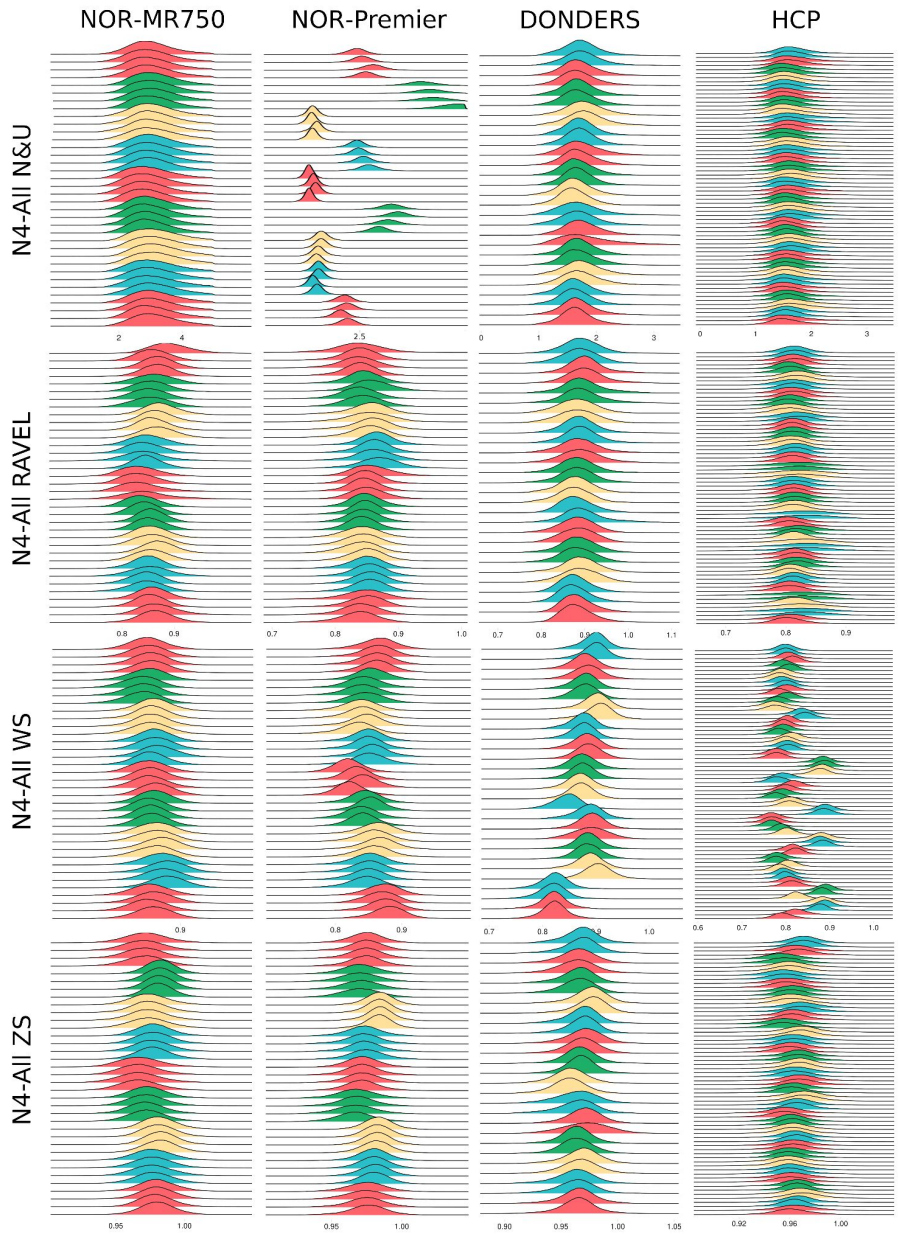
Supplementary Figure 2. T1w/T2w-ratio distributions for each participant after intensity normalisation. Adjacent distributions with the same colour represent each acquisition for that participant. Ideally, after intensity normalisation there should be minimal difference between adjacent distributions of the same colour while variation across the non-adjacent distributions and those of different colours should be retained.



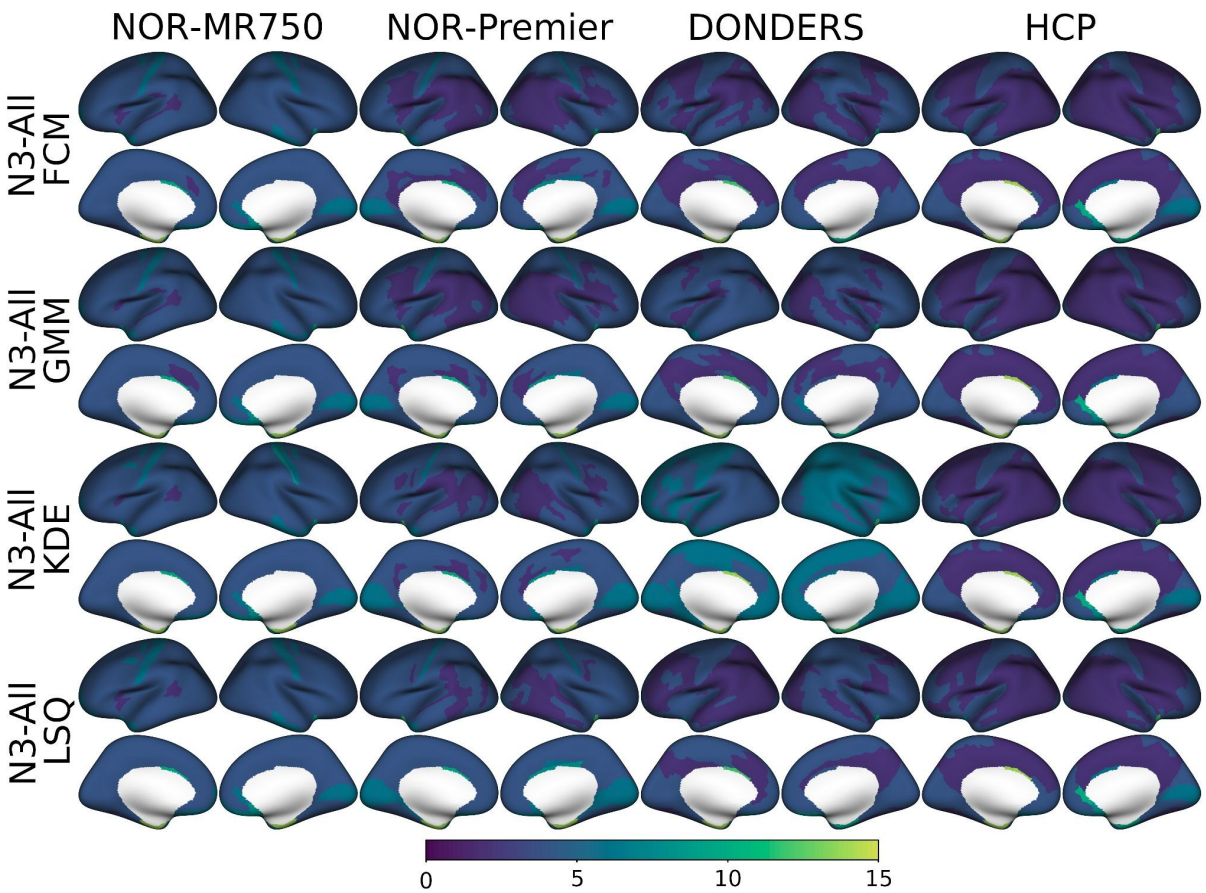
Supplementary Figure 3. T1w/T2w-ratio distributions for each participant after intensity normalisation. Adjacent distributions with the same colour represent each acquisition for that participant. Ideally, after intensity normalisation there should be minimal differencd between adjacent distributions of the same colour while variation across the non-adjacent distributions and those of different colours should be retained.



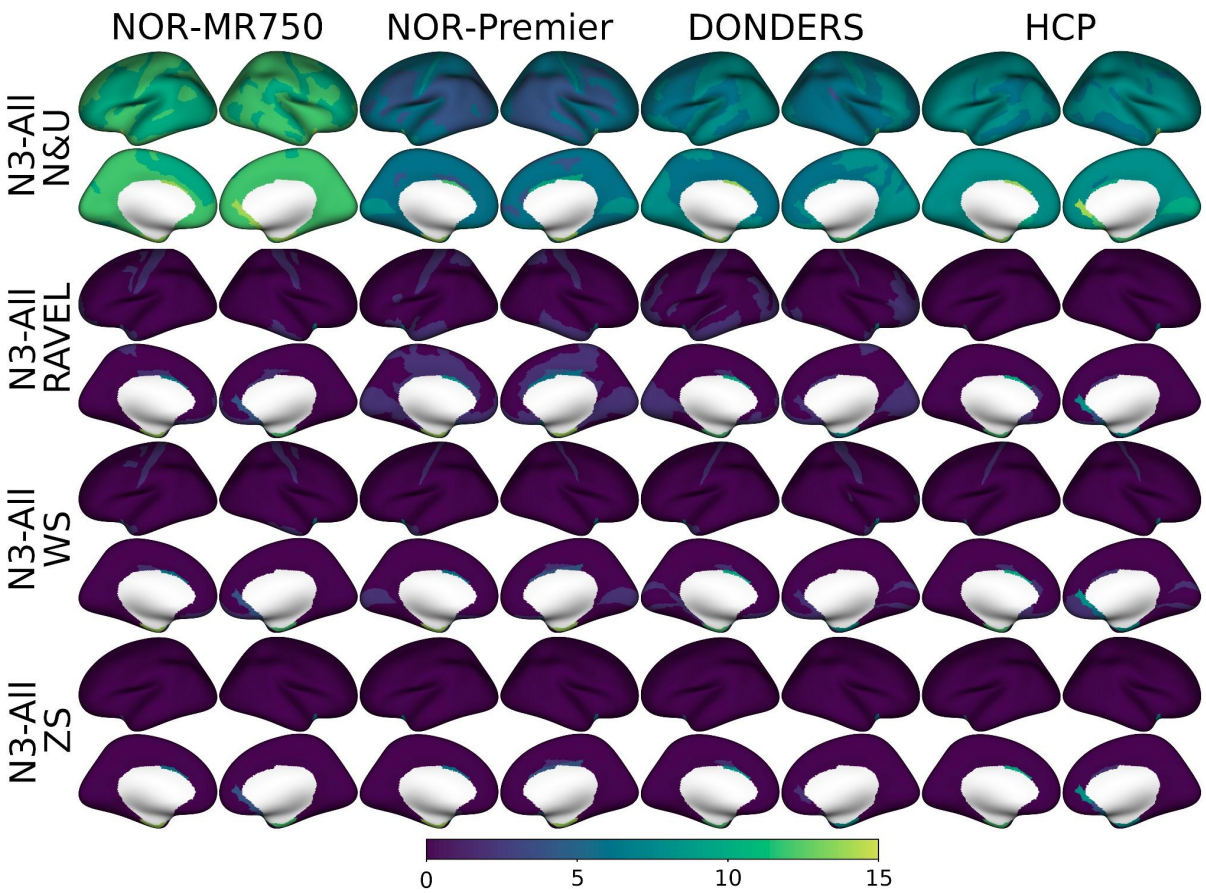
Supplementary Figure 4. T1w/T2w-ratio distributions for each participant after intensity normalisation. Adjacent distributions with the same colour represent each acquisition for that participant. Ideally, after intensity normalisation there should be minimal difference between adjacent distributions of the same colour while variation across the non-adjacent distributions and those of different colours should be retained.



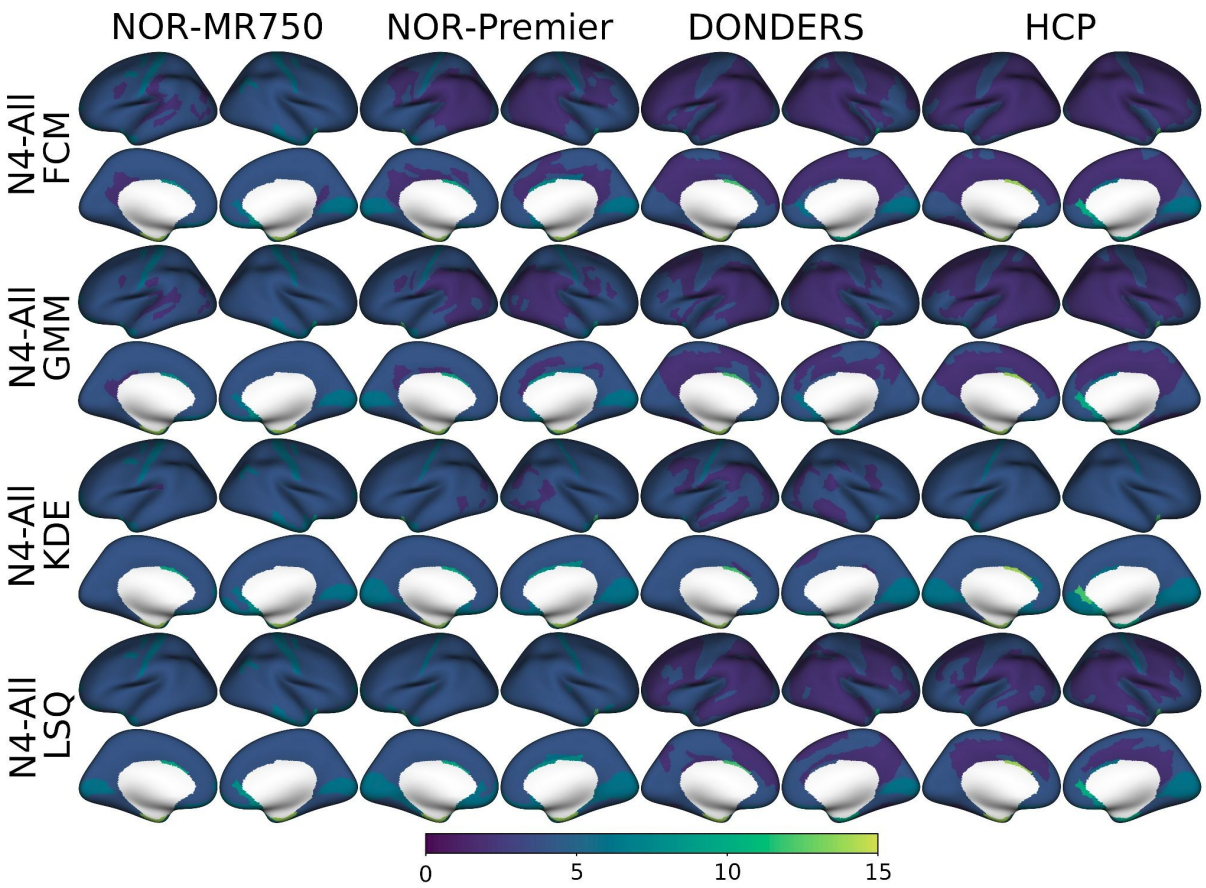
Supplementary Figure 5. T1w/T2w-ratio distributions for each participant after intensity normalisation. Adjacent distributions with the same colour represent each acquisition for that participant. Ideally, after intensity normalisation there should be minimal difference between adjacent distributions of the same colour while variation across the non-adjacent distributions and those of different colours should be retained.



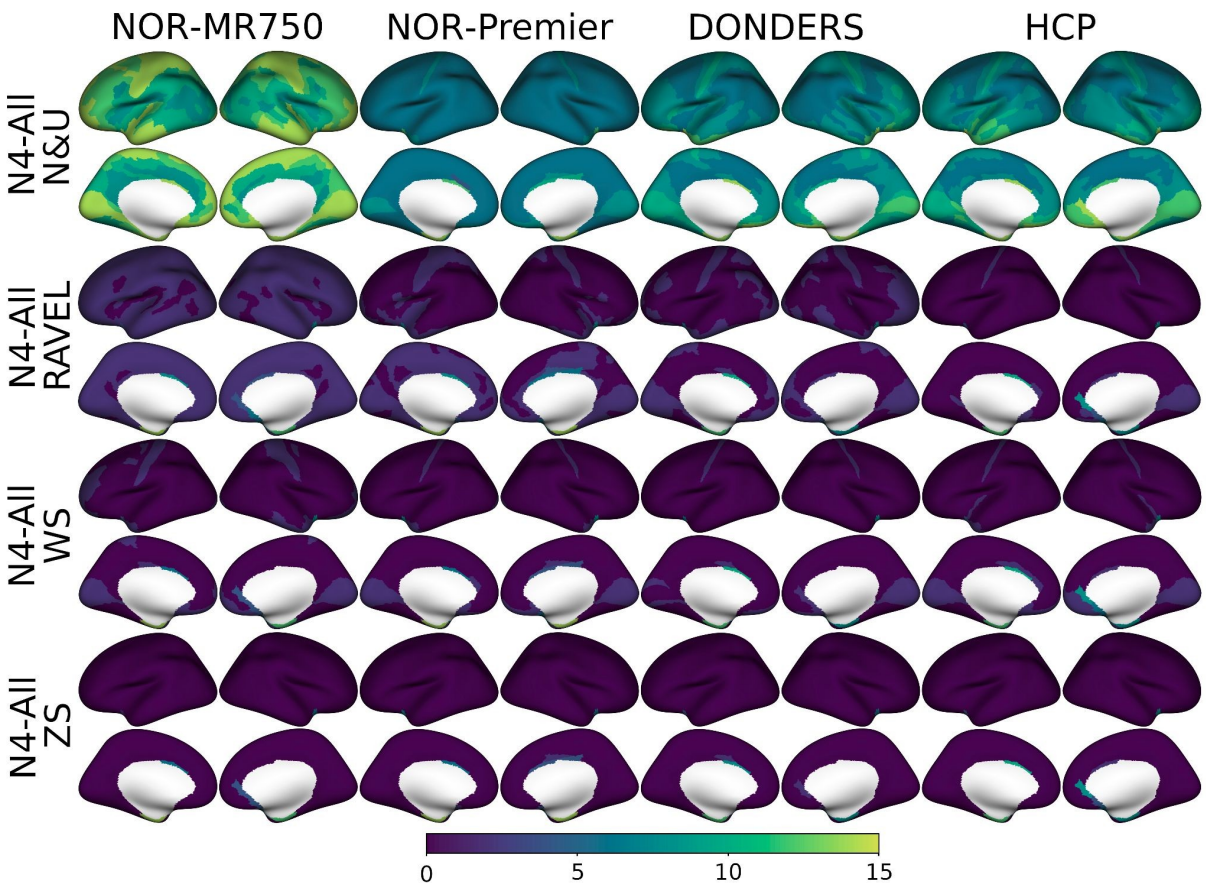
Supplementary Figure 6. CVs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



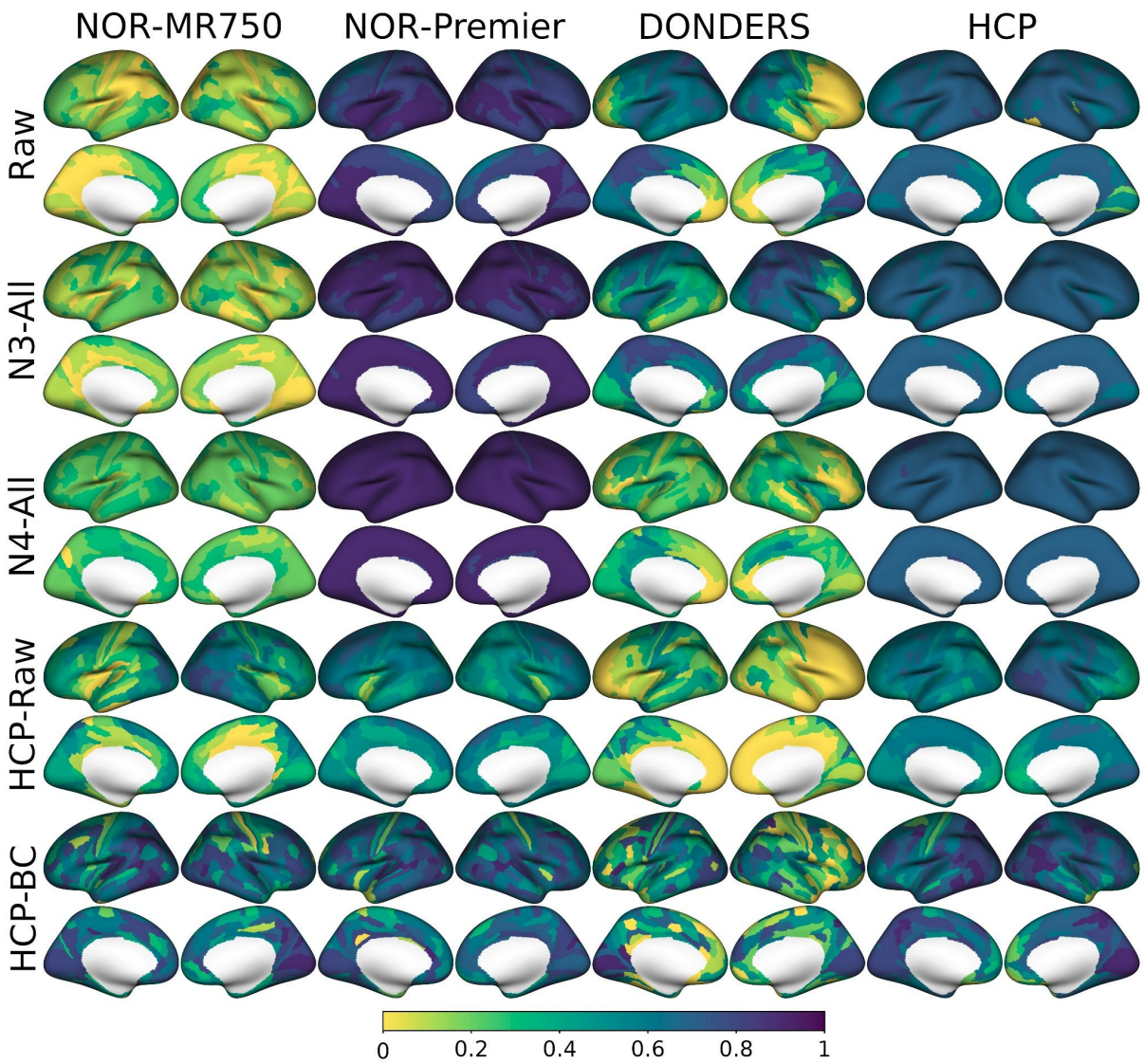
Supplementary Figure 7. CVs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



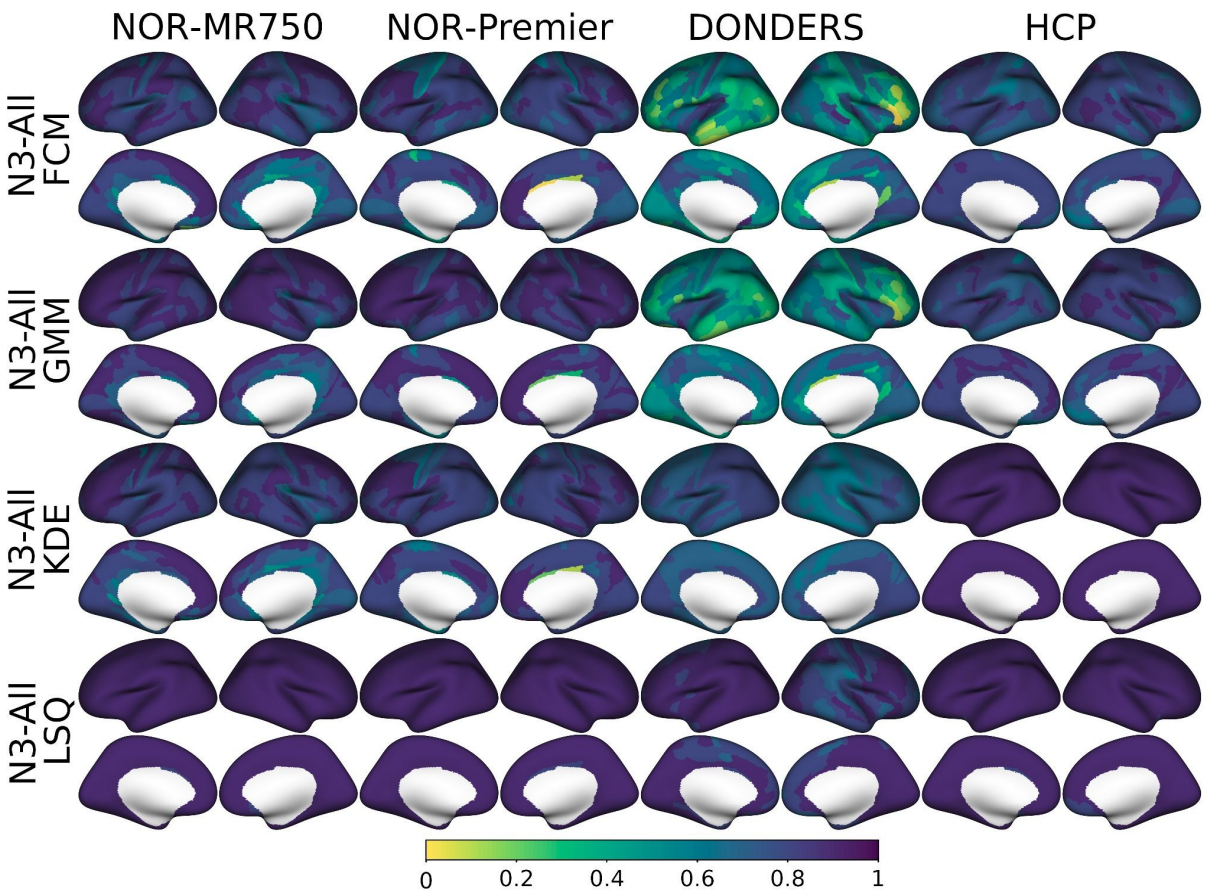
Supplementary Figure 8. CVs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



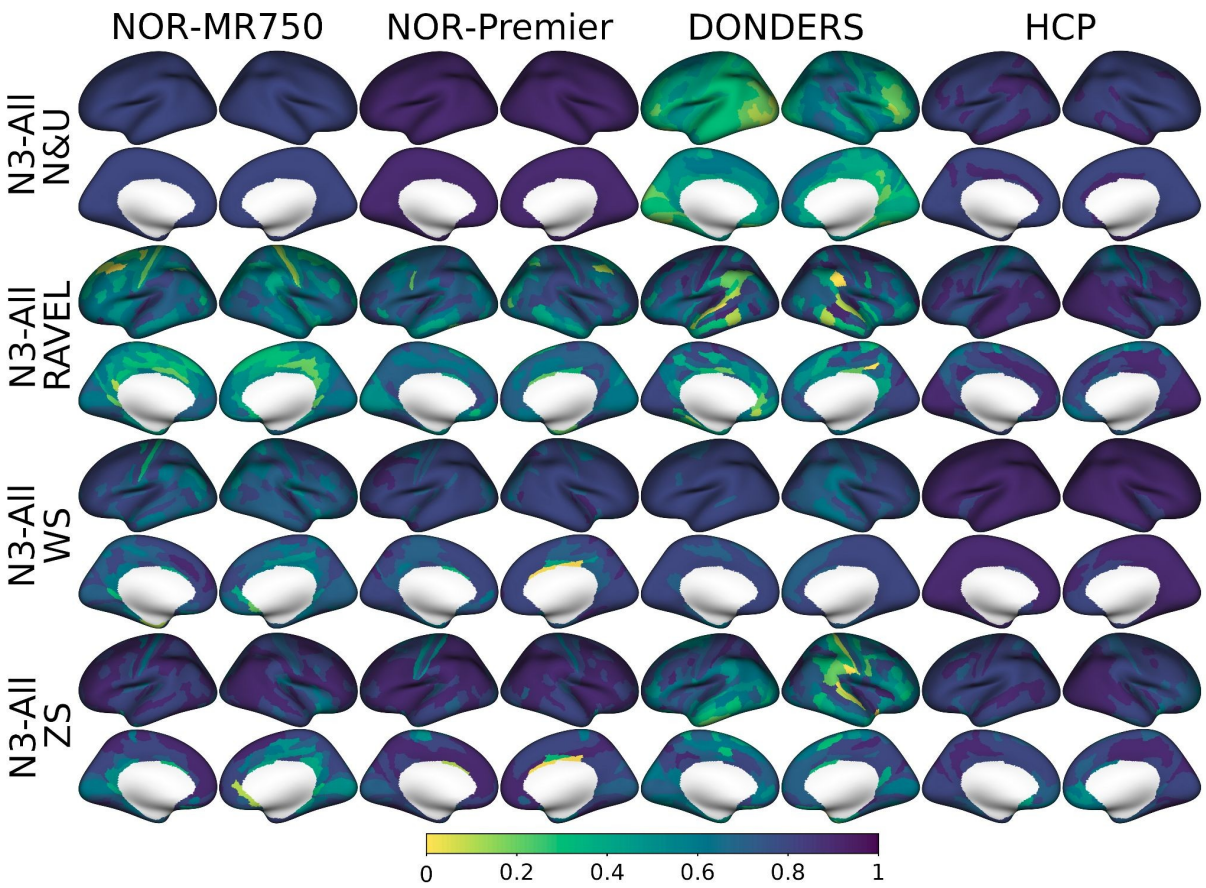
Supplementary Figure 9. CVs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



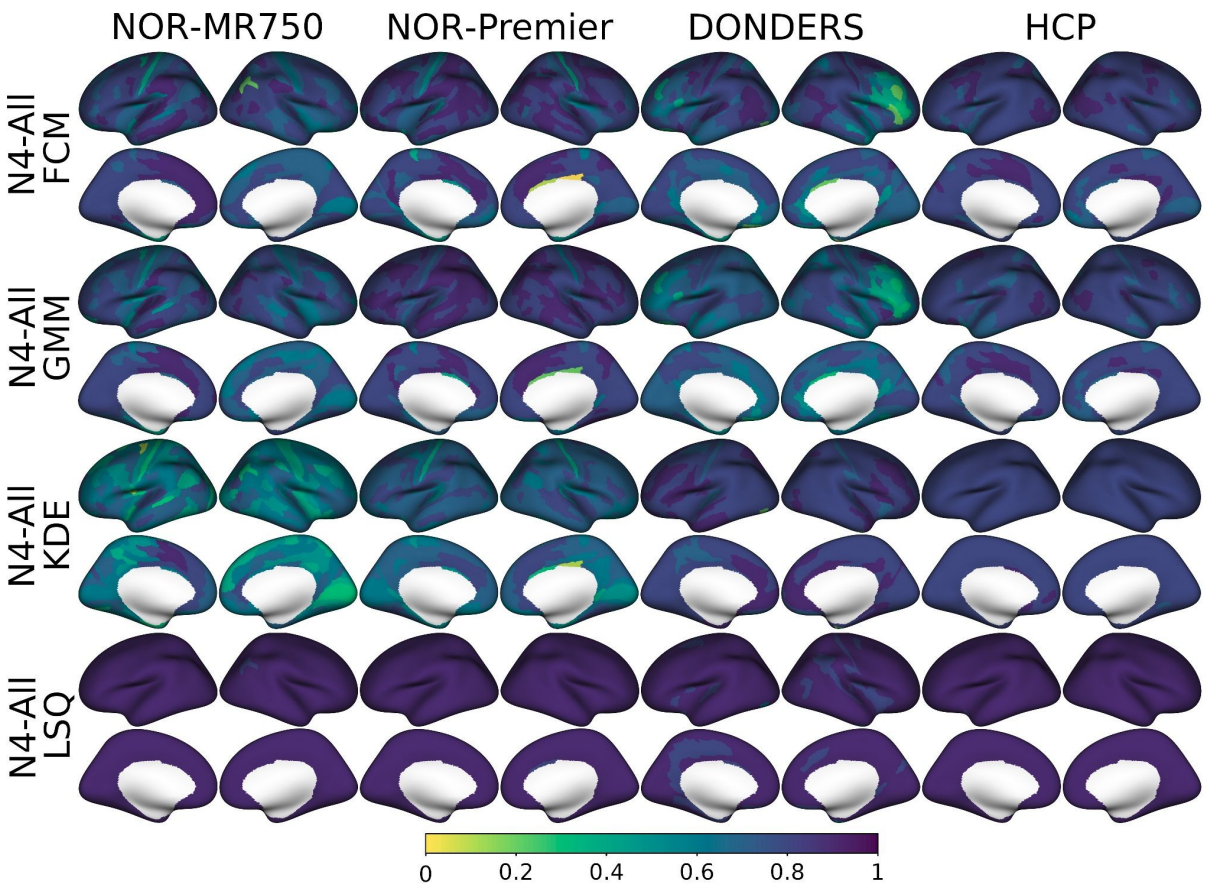
Supplementary Figure 10. ICCs in each ROI of the HCP-MMP atlas for a selection of calculation methods.



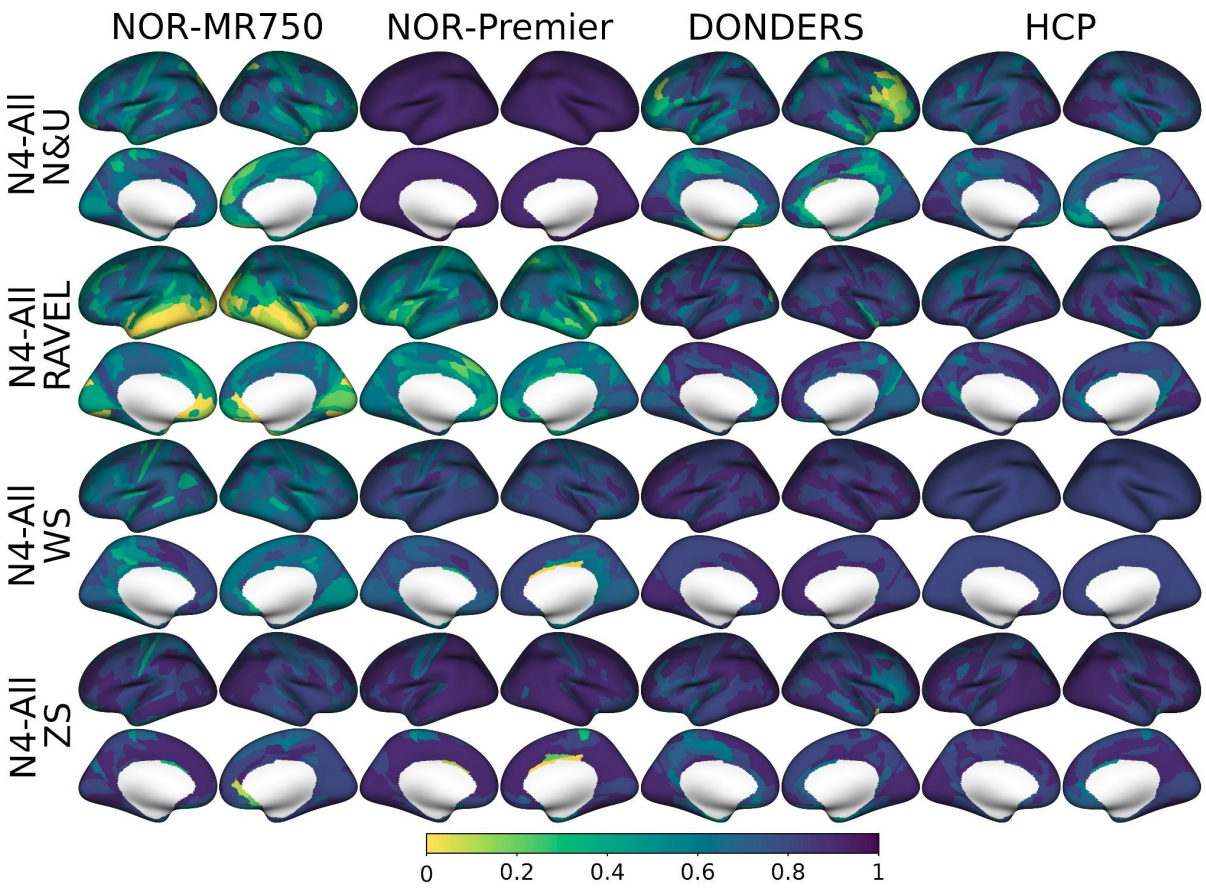
Supplementary Figure 11. ICCs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



Supplementary Figure 12. ICCs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



Supplementary Figure 13. ICCs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.



Supplementary Figure 14. ICCs in each ROI of the HCP-MMP atlas for a selection of calculation methods with intensity normalisation.

		Raw-Hires	HCP-Raw Hires	HCP-BC Hires
Mean T1w/T2w-ratio	NOR-Premier	2.02	3.20	1.29
	HCP	1.25	1.81	1.30
	Total (Mean)	1.63	2.50	1.29
Cofficient of variation (%)	NOR-Premier	6.68	5.34	13.60
	HCP	7.35	3.98	3.74
	Total (RMS)	7.02	4.71	9.98
Intraclass correlation coefficient	NOR-Premier	0.93	0.59	0.54
	HCP	0.73	0.69	0.55
	Total (Mean)	0.83	0.64	0.54
Laterality Index (%)	NOR-Premier	2.09	1.92	-0.10
	HCP	-0.16	0.14	-0.22
	Total (RMS)	1.48	1.36	0.17
Correlation with YA-BC	NOR-Premier	0.89	0.90	0.92
	HCP	0.91	0.93	0.97
	Total (Mean)	0.90	0.91	0.95
Correlation with YA-B1+	NOR-Premier	0.87	0.90	0.88
	HCP	0.76	0.77	0.88
	Total (Mean)	0.82	0.84	0.88

Supplementary Table 1. Summary statistics for the high-resolution analyses, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		HCP-Raw	HCP-BC
Correlation with YA-BC	NOR-MR750	0.22	0.93
	NOR-Premier	0.90	0.92
	DONDERS	0.79	0.96
	HCP	0.94	0.99
	Total (Mean)	0.71	0.95
Correlation with YA-B1+	NOR-MR750	-0.05	0.77
	NOR-Premier	0.91	0.89
	DONDERS	0.97	0.91
	HCP	0.77	0.88
	Total (Mean)	0.65	0.86

Supplementary Table 2. Spearman rank correlations with the reference datasets for the HCP-MPP processing methods T1w/T2w$_{HCP\text{-}Raw}$ and T1w/T2w$_{HCP\text{-}BC}$, where ROI-wise data was extracted directly from the *FS_LR* surface rather than being transformed to *fsaverage* prior to extraction.

II

III