BRAIN CORRELATES OF VULNERABILITY TO SEVERE MENTAL ILLNESS

by

Vladislav Drobinin

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
September 2020

Table of Contents

List of Tables . . . . . **vi**

List of Figures . . . . . **vii**

Abstract . . . . . **viii**

Acknowledgements . . . . . **ix**

Chapter 1 Introduction **1**

1.1 Overview 2

1.2 Risk for severe mental illness 2
1.2.1 Family history 2
1.2.2 Early antecedents 3

1.3 Brain correlates of SMI 4
1.3.1 Major depressive disorder 4
1.3.2 Bipolar disorder 5
1.3.3 Schizophrenia 6
1.3.4 Transdiagnostic features of SMI 8

1.4 Brain development during adolescence 8
1.4.1 Typical brain development 9
1.4.2 Reliability of brain data 10

1.5 Research objectives 11

Chapter 2 Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders 12

2.1 Abstract 13

2.2 Introduction 14

2.3 Methods 16
2.3.1 Participants 16
2.3.2 Materials 17
2.3.3 MRI acquisition 17
2.3.4 MRI analysis 18
2.3.5 Statistical analyses 18

2.4 Results 19

2.5 Discussion . . . 22
2.5.1 Involvement of the IFG . . . 22
2.5.2 Localization to the pars triangularis . . . 22
2.5.3 Attribution to surface area . . . 23
2.5.4 Regional increases . . . 24
2.5.5 Strengths and limitations . . . 24
2.5.6 Conclusion . . . 25

Chapter 3 Psychotic symptoms are associated with lower cortical folding in youth at risk for mental illness . . . 26

3.1 Abstract . . . 27

3.2 Introduction . . . 28

3.3 Methods . . . 30
3.3.1 Participants . . . 30
3.3.2 Participant clinical and cognitive assessment . . . 30
3.3.3 MRI acquisition . . . 31
3.3.4 MRI analysis . . . 32
3.3.5 Statistical analysis . . . 32

3.4 Results . . . 33
3.4.1 Demographic variables . . . 33
3.4.2 Overall differences in cortical folding across brain regions . . . 35
3.4.3 Prefrontal cortical folding . . . 35
3.4.4 Occipital lobe cortical folding . . . 36

3.5 Discussion . . . 37
3.5.1 Lower cortical folding . . . 37
3.5.2 Differences in prefrontal cortex folding . . . 38
3.5.3 Differences in occipital cortex folding . . . 38
3.5.4 Limitations . . . 39

3.6 Conclusion . . . 41

Chapter 4 Reliability of multimodal MRI brain measures in youth at risk for mental illness . . . 42

4.1 Abstract . . . 43

4.2 Introduction . . . 44

4.3 Materials and methods . . . 45
4.3.1 Participants . . . 45
4.3.2 Parent assessment . . . 45

4.3.3 Offspring assessment . . . 45
4.3.4 Socioeconomic status (SES) . . . 46
4.3.5 MRI acquisition . . . 46
4.3.6 MRI processing . . . 47
4.3.7 Statistical analysis . . . 48

4.4 Results . . . 49
4.4.1 Demographic and clinical characteristics . . . 49
4.4.2 Cortical volume . . . 49
4.4.3 Cortical surface area . . . 51
4.4.4 Cortical thickness . . . 51
4.4.5 Cortical folding (LGI) . . . 52
4.4.6 White matter volume . . . 52
4.4.7 Diffusion tensor imaging (DTI) measures . . . 54

4.5 Discussion . . . 55
4.5.1 Regional reliability of grey matter . . . 56
4.5.2 Regional reliability of white matter . . . 57
4.5.3 Comparisons with functional MRI . . . 58
4.5.4 Limitations . . . 59

4.6 Conclusions . . . 60

Chapter 5 Deviations from typical brain development are associated with adversity, depression, and poor functional outcomes in at-risk youth . . . 61

5.1 Abstract . . . 62

5.2 Introduction . . . 63

5.3 Methods . . . 65
5.3.1 Sample . . . 65
5.3.2 FORBOW assessments . . . 67
5.3.3 MRI acquisition . . . 68
5.3.4 MRI processing and features . . . 69
5.3.5 Statistical analysis . . . 69

5.4 Results . . . 73
5.4.1 Brain age model performance . . . 73
5.4.2 Reliability of predicted age . . . 73
5.4.3 Neuroanatomical contribution to age prediction . . . 75
5.4.4 Association with exposures and outcomes . . . 75

5.5 Discussion . . . 80
5.5.1 Limitations and future directions . . . 82

5.5.2 Conclusion 83

Chapter 6 General discussion 84

6.1 Inferior frontal gyrus in BD 84

6.2 Cortical folding and psychotic symptoms 85

6.3 Reliability of developmental MRI 85

6.4 Deviation from typical development 87

6.5 Conclusion 87

References 89

Appendix A Supplemental material 112

A.1 Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders 112

A.2 Psychotic symptoms are associated with lower cortical folding in youth at risk for mental illness 114

A.3 Reliability of multimodal MRI brain measures in youth at risk for mental illness 115

A.4 Brain Age Supplement 126

Appendix B Copyright permissions 131

List of Tables

2.1 Description of the participants 19

3.1 Demographic and clinical characteristics of the study sample . 33

5.1 Description of the cohorts 66

5.2 FORBOW Demographics 67

A.1 Grey matter volume reliability 116

A.2 Cortical surface area reliability 117

A.3 Cortical thickness reliability 118

A.4 Local gyrification index reliability 119

A.5 White matter volume reliability 120

A.6 Fractional anisotropy (FA) reliability 121

A.7 Radial diffusivity (RD) reliability 122

A.8 Mean diffusivity (MD) reliability 123

A.9 Axial diffusivity (AD) reliability 124

A.10 Reliability generalizability 14 months later 125

A.11 Scan time breakdown 126

A.12 Feature list 127

A.13 Parameter translations for xgboost model 130

A.14 Best model parameters 130

List of Figures

2.1 Larger pars triangularis volume and surface area 21

3.1 Cortical folding across the brain 34
3.2 Prefrontal cortical folding 35
3.3 Occipital cortical folding 36

4.1 Reliability of cortical gray matter measures. 50
4.2 Reliability of DTI measures. 53

5.1 Age distribution of external data 72
5.2 Brain age prediction scatterplot 74
5.3 Neuroanatomical contribution to age prediction 75
5.4 PolyE (adversity) density plot 76
5.5 YETI (antecedent) density plot 77
5.6 MDD density plot 78
5.7 CIS (functioning) density plot 79
5.8 High total loading plot 79

A.1 Structure and symptom correlations 113
A.2 Dimensional symptoms and folding 114
A.3 Variable importance plot 128
A.4 All effects forest plot 129

B.1 IFG BD licence 132
B.2 Psychotic symptoms folding licence 133
B.3 Reliability licence 134

Abstract

Severe mental illness (SMI) refers to major depressive disorder, bipolar disorder and schizophrenia. The early onset and prolonged course make SMI a leading cause of disability in the population. There is a growing need to identify biological markers of SMI to assist with early diagnosis and to inform treatment. Neuroimaging holds substantial promise in this regard. I sought to examine brain correlates of vulnerability to SMI. First, I examined SMI from a family history perspective and investigated structural changes in youth at familial risk for bipolar disorder (BD). I found that structural alterations in the inferior frontal gyrus were present not only in individuals in the early stages of BD but also in their unaffected relatives. I subsequently explored SMI from an early symptom perspective and demonstrated that attenuated psychotic symptoms during adolescence are associated with reduced cortical folding, even before the onset of psychotic illness. For biomarkers to be of clinical utility, they must be reliable. Thus, for my next project I established the scan-rescan reliability of nine commonly used structural MRI measures in our sample, including youth with anxiety and attention-deficit / hyperactivity disorder. Finally, I have compiled data from multiple developmental cohorts and built a machine learning model to quantify neuroanatomical maturity. This study shed light on how deviation in developmental trajectories relates to risk for SMI. The findings presented in my thesis contribute to a better understanding of early structural brain markers associated with risk for mental illness.

Acknowledgements

I would like to thank Rudolf Uher for his mentorship and spark for science. Sincerest thanks to my thesis committee including Chris Bowen, Matthias Schmidt, and Kazue Semba.

I would also like to thank the FORBOW team members for making the study possible, and to the FORBOW families for their generosity for volunteering their time and sharing their experiences with us. I was also fortunate to work alongside Alyson Zwicker throughout my PhD.

The presented work would not be possible without the hard work and support from Holly Van Gestel and Carl Helmick, my neuro pals.

Finally, I would like to acknowledge the tremendous support from my best friend and partner in life, Jessica Morena, throughout this degree.

These acknowledgements are inadequate to truly capture my gratitude, but thank you!

Chapter 1

Introduction

1.1 Overview

Severe mental illness (SMI) refers to functionally disruptive mental disorders that are responsible for a large proportion of disability in the population (Uher et al., 2014). SMI includes schizophrenia, bipolar disorder, and recurrent depression. SMI typically onsets in adolescence or early adulthood and often follows a life course (Caspi et al., 2020). Thus, there is great impetus to identify biological markers of SMI that may be present before illness onset. Magnetic resonance imaging (MRI) is one promising avenue for non-invasively investigating the neurobiology of psychiatric disorders. One of the strongest predictors of SMI is having a close relative who is affected (Rasic, Hajek, Alda, & Uher, 2014), inspiring this thesis to explore the relationship between familial risk and the brain. Furthermore, SMI is often preceded by developmental antecedents of mental illness (Uher et al., 2014). Therefore I will explore the relationship between early antecedents and markers of brain development. Next, this thesis will examine the reliability of commonly used structural brain measures. Finally, I will conclude by using MRI to predict an individual's stage of brain maturation and describe how deviation from the expected brain development relates to mental illness.

1.2 Risk for severe mental illness

1.2.1 Family history

One in three offspring of parents with SMI are likely to develop a major mood or psychotic disorder in the first two decades of life (Rasic et al., 2014). This estimate is higher than previously thought because prior studies have often focused on disorder-specific transmission from parent to child. Population cohorts (Dean et al., 2010), genetic evidence (Selzam, Coleman, Caspi, Moffitt, & Plomin, 2018; Uher & Zwicker, 2017), and meta-analyses (Rasic et al., 2014) have shown us that parental history of mental illness is associated with more broad transdiagnostic psychiatric outcomes in children. In other words, family history for schizophrenia elevates one's risk for psychotic disorders, but also increases risk for mood disorders, and vice versa. However, the majority of individuals who develop SMI have no apparent family history of mental illness (Al-Chalabi & Lewis, 2011), necessitating the need for supplemental risk markers.

1.2.2 Early antecedents

Longitudinal studies have shown that SMI is often preceded and predicted by developmental antecedents. Antecedents represent earlier and milder manifestations of psychopathology that may be distressing to the individual without being severely impairing. Prior literature has identified five antecedents as transdiagnostic predictors of SMI: affective lability, anxiety, sleep problems, psychotic symptoms, and basic symptoms (Uher et al., 2014).

Affective lability Affective lability corresponds to affect regulation in youth. Individuals with high affective lability have the propensity to experience strong and unpredictable changes in mood (Gerson et al., 1996). Recent work has shown that elevated affective lability is a marker of familial risk for mood disorders (Zwicker, Drobinin, et al., 2019).

Anxiety Anxiety disorders are common antecedents to SMI and precede SMI by over half a decade (Duffy, Alda, Hajek, Sherry, & Grof, 2010). Offspring of parents with mood disorders are twice as likely to have an anxiety disorder than the general population (Duffy et al., 2010; Rasic et al., 2014; Weissman et al., 2006).

Sleep problems Sleep problems early in life are another non-specific predictor of a range of physical and mental health problems later in life (Touchette et al., 2012). Abnormal sleep is also a common symptom across disorders (American Psychiatric Association, 2013).

Psychotic symptoms Psychotic symptoms, or early psychotic-like experiences, most frequently include hallucinations in the absence of a psychotic disorder. Psychotic symptoms in childhood and adolescence predict psychosis and SMI, especially if persistent rather than transitory (Arseneault et al., 2011; Poulton et al., 2000; Polanczyk et al., 2010).

Basic symptoms Basic symptoms describe subjectively perceived deficits or abnormalities in perception, cognition, language, and other important mental processes. Compared to psychotic symptoms, basic symptoms are recognized by the individual

experiencing them as atypical or unusual (Schultze-Lutter, 2009). Basic symptoms are a strong predictor of schizophrenia (Klosterkötter, Hellmich, Steinmeyer, & Schultze-Lutter, 2001; Schultze-Lutter et al., 2012).

I will examine psychotic symptoms and basic symptoms in closer detail in Chapter 3. Accumulation of several antecedents is associated with additional risk for SMI (Paccalet et al., 2016). We have developed a youth experience tracker instrument as an alternative to using multiple longer instruments to measure the aforementioned antecedents (Patterson et al., 2020). I will examine individuals with high antecedent burden in Chapter 5.

1.3 Brain correlates of SMI

Several studies suggest that alterations in brain development precede SMI onset and may help predict which youth are at risk. Despite advances in our understanding of the neurobiology of SMI, there are no MRI studies offering a complete picture of the mechanisms involved in the causes of SMI nor have neuroimaging findings made their way into informing clinical practice. Nevertheless, this section will provide a brief overview of notable studies showcasing brain alterations across SMI.

1.3.1 Major depressive disorder

A meta-analysis of the first three decades of neuroimaging in major depressive disorder (MDD) found ventricular enlargement and increased cerebrospinal fluid (CSF) volumes alongside smaller prefrontal and hippocampal volumes in MDD patients compared to controls (Kempton et al., 2011). Several recent ENIGMA (enhancing neuroimaging genetics through meta-analysis) papers have established partially overlapping findings. The examination of subcortical brain alterations revealed lower hippocampal volumes in MDD patients compared to controls (Schmaal et al., 2016). This was the only statistically significant result from one of the largest MDD studies to date, and the effect was primarily driven by patients with recurrent MDD or early age of onset.

A follow-up ENIGMA paper examined cortical alterations in adults and adolescents with MDD from 20 cohorts worldwide (Schmaal et al., 2017). Adults with MDD

had thinner cortical grey matter in the orbitofrontal cortex (OFC), anterior and posterior cingulate, as well as the insula. Adolescents with MDD overall had lower total surface area (SA) as well as regional reductions in the medial OFC and superior frontal gyrus. Lower SA was also found in sensory-motor areas. The differences in findings between adults and adolescents with MDD led the authors to suggest that MDD may affect brain structure in a dynamic way across different stages of life.

Diffusion tensor imaging (DTI) is a technique that allows for in vivo exploration of structural connectivity. Using this technique, a recent study has examined white matter microstructure alterations associated with depressive episodes during adolescence (Vulser et al., 2018). The work has revealed lower fractional anisotropy (FA) in the corpus callosum of adolescents with subthreshold depression. More interestingly, lower FA values predicted higher individual risk for depression during follow up. Finally, in the latest collaborative meta-analysis from ENIGMA, MDD was associated with advanced brain aging (Han et al., 2020). We will explore the multivariate brain age framework in Chapter 5 of this thesis, with application in MDD and beyond.

1.3.2 Bipolar disorder

Similar to MDD, one of the top structural brain changes associated with bipolar disorder (BD) is grey matter loss in the hippocampus (Cao et al., 2016). Smaller hippocampal volume was also related to worse memory performance in individuals with BD who have experienced more manic episodes. These findings were corroborated and expanded upon the same year, with the largest investigation of subcortical abnormalities in BD to date (Hibar et al., 2016). Patients with BD showed volumetric reductions in the hippocampus and thalamus coupled with enlarged ventricles. Sufficient sample sizes allowed for the comparison of BD type I and type II, however the study did not detect any significant volumetric differences between the subtypes. Lithium appeared to have a neuroprotective effect, especially on thalamic volumes. The neuroprotective effects of lithium were also found recently on a multivariate measure of brain structure (Van Gestel et al., 2019), emphasizing the importance of taking into account medication use in neuroimaging studies of psychiatric illness.

Cortical alterations have also been widely examined, including more recently in the most well-powered study of BD to date (Hibar et al., 2018). Matching prior literature,

BD was associated with lower cortical grey matter across the frontal, temporal, and parietal lobes. The strongest effects were found in the inferior frontal gyrus (IFG), the fusiform gyrus, and rostral middle frontal cortex. This report overlaps with prior findings of familial predisposition for bipolar disorder being associated with structural changes in the IFG (Hajek, Cullis, et al., 2013). I will examine this evidence in further detail throughout Chapter 2.

White matter abnormalities are also assumed to be related to the pathophysiology of BD. A recent meta-analysis across 26 cohorts has reported widespread microstructural abnormalities in BD (Favre et al., 2019). The findings revealed lower FA in patients with BD compared to healthy controls in 29 out of 43 regions examined. The highest effect sizes were reported for the corpus callosum (CC) and the cingulum. As mentioned before, reduced FA within the CC was previously found in subthreshold depression in adolescence and meta-analysis of MDD (Vulser et al., 2018; Wise et al., 2016). As we will touch upon in a moment, reduced FA within the CC has also been implicated in a meta-analysis of schizophrenia (Kelly et al., 2017), continuing a trend of shared brain alterations across SMI.

Finally, there is growing interest to move beyond mass-univariate, structure by structure, methods of finding group differences between patients and controls. Novel techniques, such as statistical / machine learning (ML), are able to capture multivariate alterations distributed throughout the whole brain. This approach may better account for the widespread yet weak effects of brain changes found in SMI. To illustrate an example, a recent study used structural MRI to identify BD in a sample of 3020 individuals (Nunes et al., 2018). The authors' algorithm was able to differentiate BD participants from controls with an accuracy of over 65%. This level of performance was well above chance, but also well below the threshold of clinical utility – an apt description of the promising yet nascent literature applying ML in psychiatry.

1.3.3 Schizophrenia

Compared to healthy controls, patients with schizophrenia exhibit smaller hippocampus, amygdala, thalamus, and intracranial volumes (van Erp et al., 2016). Indeed, reductions in whole-brain volumes have been reported in both first-episode and chronic schizophrenia patients, and are pronounced in those with a poor outcome (Haijma et

al., 2013). In Chapter 3 I will similarly note reduced intracranial volume in youth with psychotic symptoms.

The global reductions extend to measures of cortical thickness and surface area. A 39 centre worldwide collaboration reported that individuals with schizophrenia have an overall thinner cortex, and to a lesser degree, a smaller surface area than in comparable controls (van Erp et al., 2018). The largest effect sizes were observed in frontal and temporal lobes. Regional specificity was reported for cortical thickness, with the strongest bilateral reductions in the fusiform, inferior temporal, and parahippocampal gyri. Prior longitudinal work has shown greater reduction in cortical thickness among youth at clinical high risk who go on to develop psychosis (Cannon et al., 2015). Similarly, using ML techniques to assess neuroanatomical maturity, Chung et al., 2018, have shown that greater brain age deviation in youth was associated with a higher risk for developing psychosis.

Furthermore, widespread reductions have been observed in another important neurodevelopmental measure in schizophrenia. Several studies have found lower cortical folding in individuals with psychotic disorders (Nanda et al., 2014; Nesvåg et al., 2014). Aberrant gyrification has also been reported in people at genetic risk for schizophrenia (Nanda et al., 2014; Falkai et al., 2007; Liu et al., 2017) and found to be a predictor of poor treatment response in first-episode psychosis (Palaniyappan et al., 2013). In Chapter 3 I will explore cortical folding in youth experiencing psychotic symptoms.

Schizophrenia has been conceptualized as a disorder of global structural disconnectivity (Friston, Brown, Siemerkus, & Stephan, 2016). Hence neuroimaging work in schizophrenia has been interested in the integrity of white matter bundles interconnecting the cortex. The largest study of white matter microstructure in schizophrenia has found widespread reductions in FA, an indirect proxy for white matter integrity (Kelly et al., 2017). FA was lower for patients across the whole-brain white matter skeleton. Regional differences were accentuated in peripheral areas of the tract based atlas. The largest effect sizes were observed in the anterior corona radiata and corpus callosum, again partially reminiscent of the prior findings discussed in the context of mood disorders.

1.3.4 Transdiagnostic features of SMI

While summarizing the neuroimaging research I noted structural brain alterations that are common across broad diagnostic families of disorders. For example, strong effect sizes in the hippocampus, the corpus callosum, the prefrontal cortex across the SMI spectrum, despite each disorder having been analyzed independently in reference to healthy controls.

I am not the first to make this observation. In a meta-analysis of 193 studies, Goodkind et al., 2015, argued for shared neural substrates across diverse forms of psychopathology. This line of thinking has led to the re-examination of prior work and the search for transdiagnostic as well as disorder-specific structural abnormalities (Opel et al., 2020). At the same time, others have shown compelling evidence that a pervasively thinner neocortex is a transdiagnostic feature of general psychopathology (Romer et al., 2020). Romer et al., 2020, go on to argue that structural correlates of mental disorders may not follow traditional diagnostic boundaries. In addition, the authors warn against the pursuit of such specific correlates, as the endeavour might limit progress towards more effective strategies for inference, prevention, and intervention.

Whichever view is correct, neuroimaging will benefit from more case-to-case, rather than case-to-control, comparisons in the search for biomarkers across the life span.

1.4 Brain development during adolescence

Most psychiatric disorders onset in the first two decades of life (Kessler et al., 2005; Caspi et al., 2020). In attempting to understand why many psychiatric disorders emerge during adolescence, Paus, Keshavan, & Giedd, 2008, conceptualize the answer as "moving parts get broken". In other words, adolescence is characterized by major changes in the brain, and during this vulnerable period, exaggerations or abnormalities in brain maturation or environmental insults might leave a lasting mark. Thus, we should turn our attention to the 'moving parts' – or typical brain development – to ground ourselves for the rest of the thesis, which focuses on how they might 'get broken'.

1.4.1 Typical brain development

Subcortical development

A longitudinal study examined subcortical volume change across ages 8 to 22 (Tamnes et al., 2013). The **caudate, nucleus accumbens, putamen, cerebellum, and thalamus** showed the steepest volume declines during adolescence, followed by the **pallidum, hippocampus, and amygdala** with lesser or non-significant declines. In the opposite direction, the **brainstem and lateral ventricles** showed significant increases throughout development.

A more recent longitudinal study examined datasets spanning three countries to take a deeper look at the development of subcortical volumes (Herting et al., 2018). Their account was somewhat discrepant, with observed increases in **pallidum, hippocampus, and amygdala** volumes in the same 8 to 22 year old age range. The study identified sex differences in developmental trajectories. In addition, sample origin made a difference in the analysis, indicating that caution is warranted if attempting to generalize subcortical findings from data based on a single cohort.

Gross cortical development

Mills et al., 2016, investigated structural brain development between childhood and adulthood using data from four longitudinal samples. The authors found that **whole brain volume**, a combined measure of grey and white matter including the cerebellum, showed a non-linear decrease across adolescence, beginning to stabilize in the early twenties. On the other hand, **intracranial volume**, a relatively crude but widely used measure based on FreeSurfer's registration scaling, showed an annual increase of approximately 1%, beginning to stabilize in adolescence. **Cortical grey matter volume** was the highest at eight years old, the youngest age investigated. The authors noted an overall non-linear decline beginning at the youngest age, continuing to decrease through the twenties and beginning to stabilize in the thirties. Finally, **cerebral white matter volume** showed an increase between late childhood to mid adolescence. In this instance, there was some inter-sample heterogeneity in terms of when white matter volumes began to stabilize; mid or late adolescence. Furthermore, the authors observed geographical differences, as white matter volumes

estimated from youth in the United states tended to be smaller than those estimated from European samples.

Cortical surface area and thickness

Cortical grey matter volume is a product of **cortical surface area (SA)** and **cortical thickness (CT)**. SA and CT have been shown to be genetically distinct (Panizzon et al., 2009) and influenced by different neurobiological mechanisms (Winkler et al., 2010). Building on the work done by Mills et al., 2016, a follow-up study examined the individual trajectories of SA and CT in the same four site developmental sample (Tamnes et al., 2017).

The authors re-established the finding of cortical volume decreases across the sample age range (7 to 30 years old), noting a slightly accelerated decrease during adolescence. **CT** showed a non-linear decrease throughout the age range. This was in line with similar work on developmental trajectories (Walhovd, Fjell, Giedd, Dale, & Brown, 2017), in which the authors found continuous thickness decrease from 3 years onwards using multiple methods. Furthermore, the CT decrease was more prominent from ages 10 to 20, beginning to stabilize henceforth.

Tamnes et al., 2017, also reported more monotonic and steady decreases in **SA** with age, corroborating prior results (Wierenga, Langen, Oranje, & Durston, 2014). Thus the authors concluded that the largest contributor to cortical volume reductions during adolescence is reduction in CT.

1.4.2 Reliability of brain data

Accurate accounts of typical or atypical brain development depend on the reliability of the underlying neuroimaging data. Image artifcats can undermine the validity of the interpretations we make from MRI data. For example, the effects of head motion can resemble cortical grey matter atrophy on processed images (Reuter et al., 2015). This can inadvertently affect the inferences made regarding developmental trajectories (Ducharme et al., 2016). Reliability of common MRI measures has been assessed in adults (Iscan et al., 2015), however such studies are perhaps in greater need for developmental samples as younger individuals might have difficulty remaining still for the duration of MRI scans. In Chapter 4 I assess the reliability of widely used

MRI measures in youth at risk for SMI.

1.5 Research objectives

In the current thesis, I address the following research objectives:

1. Examine the relationship between family history of SMI, specifically mood disorders, and changes in the prefrontal cortex.
2. Examine the relationship between developmental antecedents of SMI, specifically basic and psychotic symptoms, and cortical folding in youth.
3. Establish the reliability of structural MRI data in youth at risk for SMI.
4. Identify how risk factors associated with mental illness relate to individual deviation from typical brain maturation.

Chapter 2

Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders

Copyright statement This chapter is based on a manuscript that has been published: Vladislav Drobinin, Claire Slaney, Julie Garnham, Lukas Propper, Rudolf Uher, Martin Alda, and Tomas Hajek. Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders (2019). *Psychological Medicine*. Re-use is permitted with copyright permission (Appendix B.1).

Contribution statement I drafted the manuscript used as the basis for this chapter, with guidance and editing from Dr. Rudolf Uher, Dr. Tomas Hajek and the other co-authors. Data were collected by the Offspring Risk for BD Imaging Study–ORBIS of which I was a part. I completed the data analysis.

2.1 Abstract

Background Larger grey matter volume of the inferior frontal gyrus (IFG) is among the most replicated biomarkers of genetic risk for bipolar disorders (BD). However, the IFG is a heterogeneous prefrontal region, and volumetric findings can be attributable to changes in cortical thickness (CT), surface area (SA) or gyrification. Here, we investigated the morphometry of IFG in participants at genetic risk for BD

Methods We quantified the IFG cortical grey matter volume in 29 affected, 32 unaffected relatives of BD probands, and 42 controls. We then examined SA, CT, and cortical folding in subregions of the IFG.

Results We found volumetric group differences in the right IFG, with the largest volumes in unaffected high-risk and smallest in control participants ($F_{2,192} = 3.07$, $p = 0.01$). The volume alterations were localized to the pars triangularis of the IFG ($F_{2,97} = 4.05$, $p = 0.02$), with no differences in pars opercularis or pars orbitalis. Pars triangularis volume was highly correlated with its SA [Pearson $r(101) = 0.88$, $p < 0.001$], which significantly differed between the groups ($F_{2,97} = 4.45$, $p = 0.01$). As with volume, the mean SA of the pars triangularis was greater in unaffected (corrected $p = 0.02$) and affected relatives (corrected $p = 0.05$) compared with controls.

Conclusions These findings strengthen prior knowledge about the volumetric findings in this region and provide a new insight into the localization and topology of IFG alterations. The unique nature of rIFG morphology in BD, with larger volume and SA early in the course of illness, could have practical implications for detection of participants at risk for BD.

2.2 Introduction

Bipolar disorders (BD) typically develop in late teens or early 20s and follow a recurrent course (Judd et al., 2002). The combination of the early age of onset and life-long course make BD one of the top causes of morbidity and disability worldwide (Judd et al., 2008). While heritability estimates for BD are as high as 89% (McGuffin et al., 2003; Song et al., 2015) there are no widely accepted biological markers of the disorder and diagnosis is made based on behavioural symptoms. This in part contributes to the fact that correct diagnosis often lags behind symptom onset by up to a decade (Ghaemi, Sachs, Chiou, Pandurangi, & Goodwin, 1999 Jan-Mar; Bschor et al., 2012). Therefore, it is imperative to identify biological markers of the disorder to assist with early diagnosis and to inform treatment (Conus, Macneil, & McGorry, 2014).

Neuroimaging holds substantial promise in investigating the neurobiology of psychiatric disorders. The largest psychiatric neuroimaging study to date has found structural brain differences in frontal, temporal and parietal regions in BD patients compared with healthy controls (Hibar et al., 2017). At the same time, meta-analyses of structural cortical alterations in BD have emphasized the heterogeneity of findings across the literature (Selvaraj et al., 2012; Wise et al., 2017). The statistical heterogeneity likely reflects clinical heterogeneity, whereas brain changes in BD might not only represent the biological markers of the disorder, but also progressive effects of illness (Hajek et al., 2012; Moylan, Maes, Wray, & Berk, 2013), common comorbidities (Hajek, Calkin, Blagdon, Slaney, & Alda, 2015; Pavlova, Perlis, Alda, & Uher, 2015), and exposure to medications (Hajek et al., 2014; Hajek & Weiner, 2016). One strategy that lends itself well to isolating the biological risk factors inherent to BD is to use genetic high-risk (HR) design. This requires the study of healthy unaffected offspring of parents with BD, who are at high genetic risk but have not yet expressed any symptoms of the illness. Therefore, neurostructural alterations in HR individuals cannot be a consequence of the illness burden, comorbidities, or treatments.

In our previous replication design HR study, we found larger right inferior frontal gyrus (IFG) volumes among both unaffected HR subjects and affected familial participants across two centres (Hajek, Cullis, et al., 2013). Other studies have reported similar findings (Adler, Levine, DelBello, & Strakowski, 2005; Matsuo et al., 2012;

Sarıçiçek et al., 2015; Roberts et al., 2016), making IFG volume change the most replicated neurostructural finding in participants at genetic risk for BD. Aside from the fact that IFG structural alterations are found in unaffected relatives, they are also associated with the illness. In addition, IFG volumes are heritable (Winkler et al., 2010). Thus, volumetric changes of IFG meet criteria for a psychiatric endophenotype (Gottesman & Gould, 2003). The localization to the rIFG also has good face validity vis-à-vis its function. The right IFG is an area of the prefrontal cortex centrally involved in response inhibition (Aron, Robbins, & Poldrack, 2014). Impaired response inhibition underlies certain symptoms of mood disorders and is a candidate neurocognitive endophenotype for BD (Bora, Yucel, & Pantelis, 2009). Indeed, hypoactivations of rIFG are frequently found in BD during inhibitory control tasks, and are also present during euthymia (see Selvaraj et al., 2012; Hajek, Alda, Hajek, & Ivanoff, 2013 for review).

Considering that IFG alterations are among the strongest candidates for a neuroanatomical signature of BD, it is important to better understand the underpinnings of these changes. Cortical grey matter volume is a product of cortical surface area (SA) and cortical thickness (CT). SA and CT have been shown to be genetically distinct (Panizzon et al., 2009), influenced by different neurobiological mechanisms (Winkler et al., 2010) and sometimes affected in opposite directions (A. Lin et al., 2017). Thus, examining SA and CT differences separately might lead to the identification of phylogenetically suitable markers that offer greater insight into the neurobiology of the disorder. Evolutionarily, gyrification (cortical folding) made possible the expansion of the cortex within the constraints of the skull and can now be accurately studied in-vivo (Schaer et al., 2008). This morphological measure captures a developmental window distinct from CT (Schaer et al., 2009) and SA (Hogstrom, Westlye, Walhovd, & Fjell, 2013). Studies of gyrification have yielded mixed results in BD (McIntosh et al., 2009; Nenadic et al., 2015). Gyrification in individuals at familial risk for BD remains to be examined.

In the current HR design study, we attempted to replicate the finding of enlarged IFG in unaffected and affected offspring of BD parents relative to controls. In addition, we investigated the localization of volumetric changes within IFG and their topology, with regard to CT, SA, and gyrification.

2.3 Methods

2.3.1 Participants

Participants were recruited from an ongoing Offspring Risk for BD Imaging Study–ORBIS. We recruited offspring from families of well-characterized adult BD probands who had participated in previous genetic and HR studies (Duffy et al., 2002; Lopez de Lara et al., 2010; Hajek, Calkin, et al., 2015) in Halifax, Nova Scotia. Participants aged 15–30 were eligible for inclusion, which represents the typical range of age at onset of BD according to cross-sectional and prospective studies (Duffy, Alda, Hajek, & Grof, 2009; Ortiz et al., 2011). Consequently, these individuals remain at a substantial risk of future onset of BD. Only the offspring, not the probands, participated in the MRI study. As this is a part of an ongoing study, the sample included here was larger and partially overlapping with the sample described in our previous publication – 103 v. 82 participants (Hajek, Cullis, et al., 2013).

In keeping with previous studies (Todd et al., 1996; Duffy et al., 2002), we included participants with BD type I or type II, but not with BD NOS as probands for this study. Similar to participants with BD I, the BD II individuals had a low prevalence of comorbid conditions and an episodic course of illness. Family studies focusing on similarly narrow diagnoses have generally found them to be part of the same genetic spectrum (Gershon et al., 1982). Furthermore, neuroimaging studies show similar findings for patients with BD I and BD II (Hamakawa, Kato, Murashita, & Kato, 1998; Winsberg et al., 2000; Dager et al., 2004; McGrath, Wessels, Bell, Ulrich, & Silverstone, 2004; Silverstone, Asghar, O'Donnell, Ulrich, & Hanstock, 2004; Wu et al., 2004).

The offspring from BD probands were divided into two subgroups. (1) The unaffected HR group, which included 32 offspring without a personal history of Axis I psychiatric disorders. These individuals were considered HR because they came from multiplex families (more than one member affected with BD) and had one parent affected with a primary mood disorder. (2) The affected familial group, which included 29 offspring meeting criteria for a lifetime Axis I diagnosis of mood disorders (i.e. a personal history of at least one episode of depression, hypomania, or mania meeting full DSM-IV criteria) and had one parent affected with a primary mood disorder.

Depressive episodes were included because unipolar depression is characteristically the first manifestation of illness in patients who later develop BD (Duffy et al., 2002; Hillegers et al., 2005). Additionally, around 70% of depressed first-degree relatives of BD probands are estimated to suffer from BD (Blacker & Tsuang, 1993). Lastly, we recruited 42 control participants free of personal or family history of DSM-IV Axis I psychiatric disorders.

Common exclusion criteria for all groups were a personal history of (1) any serious medical or neurologic disorders, (2) substance abuse/dependence during the previous 6 months, or (3) magnetic resonance imaging (MRI) exclusion criteria.

2.3.2 Materials

Each participant received a complete description of the study and provided written informed consent. The study was approved by the Research Ethics Boards of the Izaak Walton Killam Health Center and the Nova Scotia Health Authority, Halifax, Nova Scotia.

Probands, offspring and control subjects were interviewed by pairs of clinicians (psychiatrists and/or nurses) using Schedule for Affective Disorders and Schizophrenia—Lifetime version (SADS-L) (Endicott & Spitzer, 1978) or Kiddie Schedule for Affective Disorders and Schizophrenia, Present and Lifetime version (KSADS-PL) (Kaufman et al., 1997) for participants under 18 years of age. Diagnoses were made based on DSM-IV in a blind consensus review, by an independent panel of senior clinical researchers using all available clinical materials.

2.3.3 MRI acquisition

MRI acquisitions were performed with a 1.5-Tesla General Electric Signa scanner equipped with a single-channel head coil, located at the IWK Health Centre, Halifax, Nova Scotia. After a localizer scan, a T_1-weighted spoiled gradient recalled (SPGR) scan was acquired with the following parameters: flip angle 40°, echo time 5 ms, repetition time 25 ms, field of view 24 cm $\times$ 18 cm, matrix 256 $\times$ 160 pixels, number of excitations = 1, no interslice gap, 124 coronal 1.5 mm thick slices.

2.3.4 MRI analysis

The structural T_1-weighted scans were analyzed using FreeSurfer v5.3 (Fischl et al., 2004). Briefly, FreeSurfer processing included automated skull stripping, bias field correction, non-linear registration with a stereotaxic atlas and grey-white matter segmentation and generation of cortical surface models. Utilizing these models, an automated labelling system subdivided the cerebral cortex into gyral-based regions of interest (ROIs) corresponding to the Desikan–Killiany Freesurfer atlas (Desikan et al., 2006). For each participant, we extracted cortical grey matter volume, thickness, and SA for three subdivisions of the bilateral IFG: pars orbitalis, pars triangularis, pars opercularis.

Furthermore, the pial surface reconstruction allowed us to measure the 3D local gyrification index (lGI). The lGI measure provides a ratio of the convex hull (outer) cortex SA to buried cortex SA. Thus, a greater lGI represents greater gyrification, or more folding in each ROI and vice versa. Details of the automated lGI computation can be found in the validation paper (Schaer et al., 2008) and on the FreeSurfer website.

Raw images were inspected for motion, low contrast, and other artifacts. FreeSurfer output was manually examined for skull stripping and segmentation errors. If found, errors were manually corrected and the affected scans re-processed.

2.3.5 Statistical analyses

Statistical analyses were performed in R Studio (R version 3.3.3). Demographic and clinical variables were compared with one-way Analysis of Variance (ANOVA) or a χ^2 test for categorical variables. For the primary analysis, we investigated group differences in bilateral IFG volume. For this, we used Multivariate Analysis of Variance (MANOVA), with the IFG volume and individual IFG subdivisions (pars orbitalis, pars triangularis, pars opercularis) as the dependent variable, group status as an independent variable, age, sex, and estimated total intracranial volume (TIV) (Buckner et al., 2004) as covariates. In follow-up one-way ANOVAs we explored the localization of volumetric changes within the IFG. Volumes of IFG subdivisions served as dependent variables with group status as the independent variable and age, sex, and TIV as covariates. We controlled for multiple comparisons with the Benjamini-Hochberg

Table 2.1: Description of the participants

	Unaffected HR	Affected Familial	Control	*p* Value
N	32	29	42	NA
Sex, *n* (%) female	23 (65.62)	19 (65.51)	25 (59.52)	ns
Age mean (s.d.)–years	19.25 (3.14)	20.91 (3.66)	21.94 (3.72)	0.014
Related to another study participant, *n* (%)	14 (43.75)	7 (24.14)	8 (19.05)	ns
Diagnosis	NA	1 ADD, 4 BDI, 3 BDII, 2 BD NOS[a], 18 MD, 1 psychosis NOS	NA	NA
Comorbid conditions	NA	1 ADHD, 4 EO, 5 ADO, 2 SUD	NA	NA
Treatment at the time of scanning, *n* (%)	NA	10 (34.48)	NA	NA
Medication type at the time of scanning	0 (0)	2 AAP, 1 AC, 5 AD, 2 Li, 1 PS	0 (0)	NA
Lifetime history of Li treatment, *n* (%)	0 (0)	5 (17.24)	0 (0)	NA
Illness duration–years, mean (s.d.)[b]	NA	4.56 (2.91)	NA	NA
No. of episodes, mean (s.d.)[b]	NA	2.56 (2.68)	NA	NA
Intracranial volume–cm^3, mean (s.d.)	1526.93 (129.38)	1522.96 (161.79)	1505.29 (144.67)	ns

AAP, atypical antipsychotics; AC, anticonvulsant; AD, antidepressants; ADD, adjustment disorder; ADHD, attention-deficit/hyperactivity disorder; ADO, anxiety disorder; BD, bipolar disorder; EO, eating disorder; HR, high risk; Li, lithium; MD, major depression; NA, not applicable; NOS, not otherwise specified; ns, nonsignificant, PS, psychostimulant; SUD substance use disorder.
[a]BD NOS participants met criteria for major depressive episodes and had subsyndromal hypomanic symptoms.
[b]Data missing in two participants.

procedure (Benjamini & Hochberg, 1995) and report corrected p values. In subregions with significant between-group volumetric differences, we further explored which topographical features most contributed to these alternations (CS, SA, gyrification). We used identical model specifications as above for these exploratory ANOVAs followed by Bonferroni corrected pair-wise comparisons using Tukey's Honest Significant Difference (HSD) test to identify which groups were driving the significant omnibus findings. We also used Pearson's correlation coefficient to quantify the strength of the relationship between regional volumetric differences and corresponding measures of SA, thickness, and gyrification.

As several participants in the study were biologically related, we did sensitivity testing of significant results using mixed effect generalized linear models structured identically to the ANOVA but with the additional inclusion of family membership as a random effect. Effect sizes are summarized with partial eta squared (η_p^2) and its bootstrapped 95% confidence interval (CI) over 1000 simulations.

2.4 Results

We recruited 32 unaffected HR, 29 affected familial, and 42 control participants. The groups did not differ in the proportion of females or intracranial volumes, but control participants were older than the unaffected HR participants (see Table 2.1).

When controlling for age and sex, MANOVA revealed that the unaffected, affected, and control participants differed in the right ($F_{2,192} = 3.07$, $p = 0.01$), but not the left ($F_{2,192} = 1.61$, $p = 0.29$), IFG volumes. The overall rIFG volume was largest in unaffected and smallest in control participants.

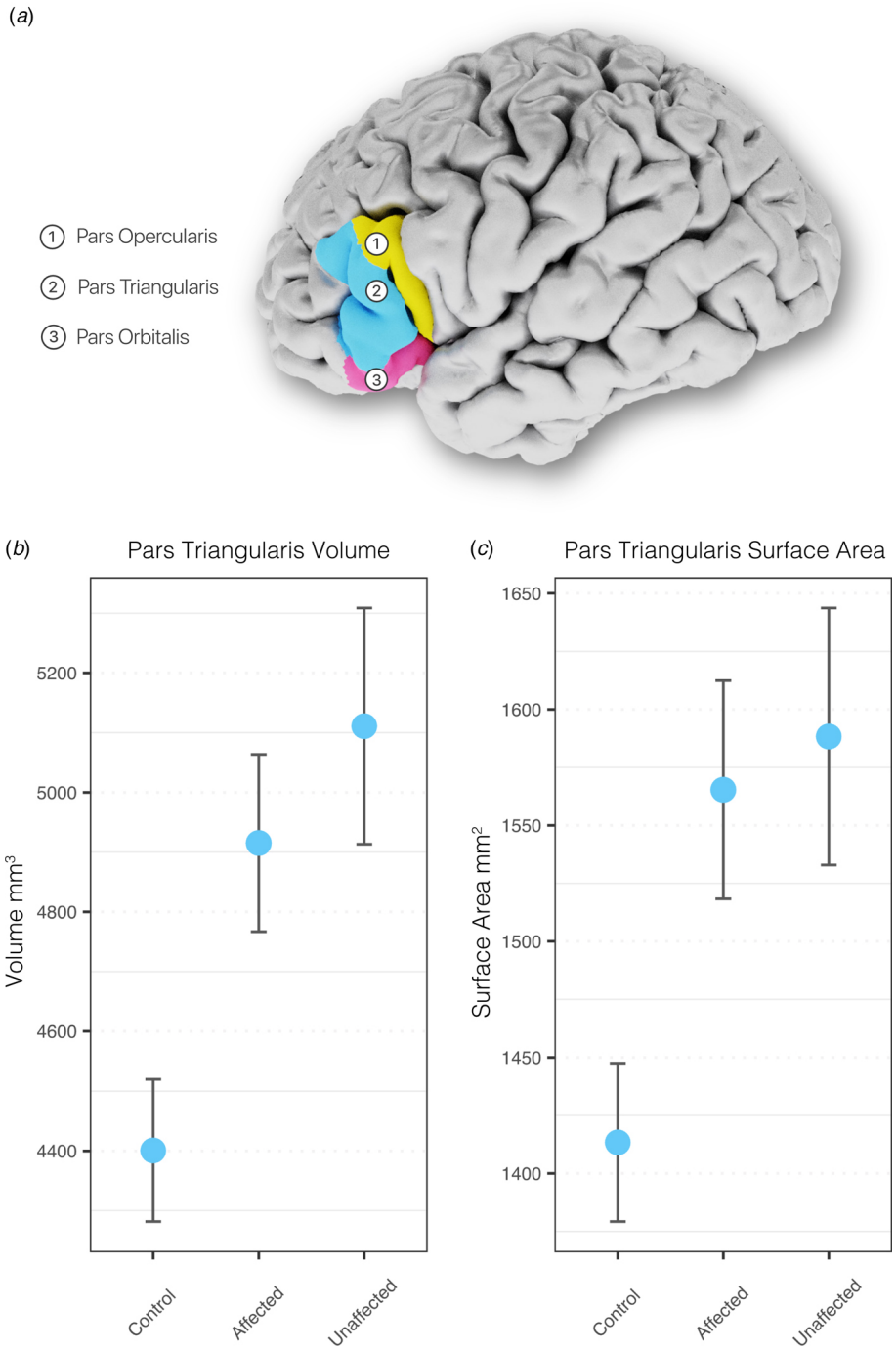
Within the right IFG, we found the largest volume differences in the pars triangularis ($F_{2,97} = 4.05$, $p = 0.02$, $\eta_p^2 = 0.08$, 95% CI 0.001–0.18), see Fig 2.1. Post hoc comparisons using the Tukey HSD test showed the volume of the right pars triangularis was greater in unaffected HR (M = 5110.88 mm^3, s.d. = 1118.43 mm^3, p = 0.02) and affected familial groups (M = 4915.17 mm^3, s.d. = 798.88 mm^3, p = 0.05) compared with control participants (M = 4400.67 mm^3, s.d. = 771.70 mm^3). The group differences in pars triangularis remained significant when we controlled for family membership ($F_{2,97} = 3.12$, $p = 0.05$). There were no significant volumetric differences in the pars opercularis ($F_{2,97} = 2.85$, $p = 0.06$, $\eta_p^2 = 0.006$, 95% CI 0–0.15) or the pars orbitalis ($F_{2,97} = 1.71$, $p = 0.2$, $\eta_p^2 = 0.03$, 95% CI 0–0.12).

The volume of the right pars triangularis was highly correlated with its SA [Pearson r(101) = 0.88, p <0.001] and ANOVA revealed significant group differences in pars triangularis SA ($F_{2,97} = 4.45$, $p = 0.01$, $\eta_p^2 = 0.08$, 95% CI 0.003–0.19). Post hoc comparisons using the Tukey HSD test showed that as with volume, the mean SA of the right pars triangularis was greater in unaffected HR (M = 1588.31 mm^2, s.d. = 313.29 mm^2, $p = 0.02$) and affected familial groups (M = 1565.38 mm^2, s.d. = 253.33 mm^2, $p = 0.03$) compared with control participants (M = 1413.38 mm^2, s.d. = 221.22 mm^2).

The correlation between the right pars triangularis grey matter volume and CT was much lower than for SA, but still significant [Pearson r(101) = 0.28, $p = 0.005$]. However, we did not find differences between the groups in CT ($F_{2,97} = 0.05$, p = 0.95, $\eta_p^2 = 0.001$, 95% CI 0–0.01) of the pars triangularis. As expected SA and CT of this region were not associated [Pearson r(101) = -0.16, $p = 0.11$].

Pars triangularis volume and gyrification were not significantly correlated [Pearson r(101) = –0.13, $p = 0.19$]. Likewise the groups did not differ in gyrification ($F_{2,97}$ = 0.06, $p = 0.94$, $\eta_p^2 = 0.001$, 95% CI 0–0.018) of the pars triangularis.

Figure 2.1: Larger pars triangularis volume and surface area in affected familial and unaffected High-Risk groups relative to control participants. Means ± s.e.m. (a) Anatomical subdivisions of the inferior frontal gyrus (IFG). Significant group differences were localized to pars triangularis. (b) Volume of pars triangularis. (c) Surface area of pars triangularis.



2.5 Discussion

In this study, we replicated and extended the finding of increased right IFG volume as a neuroanatomical marker of genetic susceptibility for BD. Using 3D representations of the cortical sheet we found that the volumetric enlargement was linked to the increased cortical SA, and not CT or cortical folding. Moreover, we localized the largest volume and SA differences to the pars triangularis of the rIFG, providing a more sensitive marker of genetic risk for BD.

2.5.1 Involvement of the IFG

Our finding of larger rIFG follows a well-established pattern of evidence implicating this region of the prefrontal cortex in BD. The direction of the finding may seem surprising, but regional cortical volume increases have been reported in other HR and early-course BD imaging studies (Kempton et al., 2009; Frangou, 2011; Adleman et al., 2012; Bauer et al., 2014; K. Lin et al., 2015). Increased IFG grey matter volume has previously been found in the early course of BD, in patients with the first episode of mania (Adler et al., 2005). Our previous exploratory, replication design, multi-centre study has shown larger rIFG volumes in unaffected HR, affected familial as well as in a third group comprising a clinical sample of young BD participants (Hajek, Cullis, et al., 2013). Similar findings were replicated in more recent studies, showing increased IFG volumes in both euthymic BD patients and healthy first-degree relatives when compared with healthy controls (Sarıçiçek et al., 2015; Roberts et al., 2016). In addition, the IFG has been shown to be sensitive to illness chronicity, with studies finding a significant negative correlation between the duration of illness and grey matter volume in this region (Matsuo et al., 2012; Hajek, Cullis, et al., 2013).

2.5.2 Localization to the pars triangularis

In this study, we found that the IFG neurostructural alterations were most pronounced in the pars triangularis, a finding in concordance with the area's functional profile. Studies in healthy individuals and lesion studies have shown that the pars triangularis of the IFG is not only associated with but also necessary for successful response inhibition (Collette et al., 2001; Aron et al., 2014; Menon, Adleman, White,

Glover, & Reiss, 2001). In fact, age-related improvements in activation of the pars triangularis within the frontostriatal network underlie improvements in the normative development of self-regulatory control (Marsh et al., 2006). Impaired response inhibition is central to many symptoms of mania, such as increased risk-taking, impulsive behaviour, talkativeness, and excessive spending. Moreover, a meta-analysis of neurocognitive studies in at-risk participants revealed response inhibition as the most prominent endophenotype of BD (Bora et al., 2009). As would be expected, BD is marked by diminished activity in the IFG, particularly during response inhibition to emotional stimuli (Hajek, Alda, et al., 2013) and during reward processing (Singh et al., 2014). Hypoactivity in the IFG has been reported in both BD patients and youth at HR for the disorder as well as during manic and euthymic states (Foland-Ross et al., 2012; Townsend et al., 2012; Roberts et al., 2013), further corroborating the endophenotypic nature of the rIFG alterations.

2.5.3 Attribution to surface area

Cortical grey matter volume is a product of CT and SA. Evidence suggests that CT is determined by asymmetric division of radial glia progenitors and reflects the number of cells within cortical columns, while SA is linked to symmetric division of progenitor cells in the ventricular and sub ventricular layers and relates to the quantity of cortical columns (Rakic, Ayoub, Breunig, & Dominguez, 2009; Florio & Huttner, 2014; Wierenga et al., 2014). Thus, the two measures develop with distinct cellular mechanisms and show almost no genetic correlation (Panizzon et al., 2009). The finding that increased pars triangularis volume was driven by SA and not CT is in line with the general finding that individual variation in human cortical volume is more attributable to variation in SA than CT (Im et al., 2008). Furthermore, the SA of the right pars triangularis is significantly more heritable than its CT (Winkler et al., 2010). Our finding is further supported by a recent BD twin-study showing that genes influencing BD are associated with regional increases in SA (Bootsman et al., 2015). Accordingly, a recent pilot study utilized regional increases in SA to distinguish BD from major depressive disorders (Fung et al., 2015).

2.5.4 Regional increases

Regional increases in brain volume or SA have been attributed to resilience (Ladouceur et al., 2008), neuropathology (Adler et al., 2005), effects of medication (Yucel et al., 2007; Hajek et al., 2014), or disrupted maturation (Konarski et al., 2008). The larger pars triangularis volume and SA is unlikely to be a marker of resilience as it was also larger in already affected participants. Furthermore, resilience would be difficult to infer from the current status of unaffected HR participants as they are passing through the HR age range and can develop BD in the future. Neuropathology, such as hypertrophy due to preapoptotic edema is doubtful since our previous spectroscopy work with a partially overlapping sample showed similar metabolite concentrations surrounding the IFG between unaffected HR and control participants (Hajek et al., 2008). Even more importantly, edema would likely lead to changes in both CT and SA. Neurotrophic effects of medication, such as lithium, need to be considered. However, in our sample, the largest structural differences were found in the unaffected HR participants who were medication naive. Normative brain maturation from childhood to young adulthood involves grey matter reductions across the cortex. It has been suggested that increased grey matter volumes in BD offspring might be indicative of disruptions in normative brain growth, for example, through an aberrant synaptic remodeling (Sowell et al., 2003; Herting, Gautam, Spielberg, Dahl, & Sowell, 2015). However, while disrupted maturation is plausible, we controlled for age and prospective studies would be more suitable in inferring developmental trajectories.

2.5.5 Strengths and limitations

The strengths of this study are its hypothesis-driven approach, young sample, and genetic HR design. We have conducted a targeted investigation of the IFG because of its strong previously established involvement in BD. Furthermore, we investigated this ROI in a relatively large dataset of 103 brain scans. The participants were mostly in their 20s, recruited from the age range during which transition to BD approaches peak incidence. We benefitted from also scanning unaffected offspring of BD probands because neuroanatomical changes have previously been linked to confounds such as illness burden, common comorbidities, and effects of treatment, none of which apply to the unaffected group.

There are several limitations of this study. While we characterized the structural phenotype and topology of IFG alterations in BD, we did not include measures of cognitive functioning or genetic markers. Future work should investigate the relationship between pars triangularis structural change with neurocognitive test performance. In addition, genotyping data would allow future studies to test whether the association between IFG structural change is moderated by a polygenic risk score for BD. Furthermore, our study was not designed to capture the effects of sub-syndromal features. There was no association between cortical measures and clinical scales (Supplement Figure A.1).

Our hypothesis-driven analysis was focused on the IFG, however, this is not the only region involved in the disorder (Hibar et al., 2018) and BD biomarkers will be further enhanced by incorporating information from additional brain regions and their interactions, in a multivariate statistical-learning framework. Interestingly, previous machine-learning work has also supported the importance of IFG structure in identifying participants at risk for BD (Hajek, Cooke, et al., 2015; Roberts et al., 2017).

Currently, it is unknown if the pattern of SA increases without CT changes is dynamic through the development and course of illness. Future prospective longitudinal studies can provide additional context into the developmental window of BD. Finally, a proportion of our affected familial and unaffected HR participants were related thus introducing non-independence of cortical structure. Our main findings remained unchanged when we controlled for this by implementing mixed effects models with family membership as a random effect.

2.5.6 Conclusion

In summary, we expanded the finding of increased rIFG volume as a marker of genetic risk for BD. We localized the largest group differences to the enlargement of the right pars triangularis and found that volumetric change was driven mainly by SA rather than CT or folding. These findings strengthen prior knowledge about the volumetric alterations in this region and provide new insight into the localization and topology of IFG alterations, which aid in better understanding of brain risk factors associated with BD.

Chapter 3

Psychotic symptoms are associated with lower cortical folding in youth at risk for mental illness

Contribution statement I drafted the manuscript, completed the data analysis, and devised the original idea for the chapter. I received guidance and editing from Dr. Rudolf Uher, and the other co-authors. Data were collected by the FORBOW research team. I collected data and assisted with the ongoing training of neuroimaging staff and students.

3.1 Abstract

Background Cortical folding is essential for healthy brain development. Previous studies have found regional reductions in cortical folding in adult patients with psychotic illness. It is unknown whether these neuroanatomical markers are present in youth with subclinical psychotic symptoms.

Methods We collected MRIs and examined the local gyrification index in a sample of 110 youth (mean age ± standard deviation 14.0 ± 3.7 yr; range 9–25 yr) with a family history of severe mental illness: 48 with psychotic symptoms and 62 without. Images were processed using the Human Connectome Pipeline and FreeSurfer. We tested for group differences in local gyrification index using mixed-effects generalized linear models controlling for age, sex and familial clustering. Sensitivity analysis further controlled for intracranial volume, IQ, and stimulant and cannabis use.

Results Youth with psychotic symptoms displayed an overall trend toward lower cortical folding across all brain regions. After adjusting for multiple comparisons and confounders, regional reductions were localized to the frontal and occipital lobes. Specifically, the medial (β = –0.42, p_{FDR} = 0.04) and lateral (β = –0.39, p_{FDR} = 0.04) orbitofrontal cortices as well as the cuneus (β = –0.47, p_{FDR} = 0.03) and the pericalcarine (β = –0.45, p_{FDR} = 0.03) and lingual (β = –0.38, p_{FDR} = 0.04) gyri.

Limitations Inference about developmental trajectories was limited by the cross-sectional data.

Conclusion Psychotic symptoms in youth are associated with cortical folding deficits, even in the absence of psychotic illness. The current study helps clarify the neurodevelopmental basis of psychosis at an early stage, before medication, drug use and other confounds have had a persistent effect on the brain.

3.2 Introduction

Psychosis is marked by hallucinations, delusions and disturbances of affect and behaviour. It manifests in major depressive disorder, bipolar disorder and most notably, schizophrenia. Although psychotic disorders are a leading cause of morbidity and disability worldwide (Global Burden of Disease Study 2013 Collaborators, 2015), their etiology remains unclear (Radua et al., 2018). Multiple lines of evidence are converging to suggest that the disease stems from processes that affect neurodevelopment (Rapoport, Giedd, & Gogtay, 2012). The neurodevelopmental hypothesis posits that genetic and environmental risk factors perturb early brain development, leading to symptoms later in life as the brain matures and copes with new stressors (Selemon & Zecevic, 2015). One process necessary for healthy brain development is cortical folding, which may be abnormal in the pathogenesis of schizophrenia (White & Gottesman, 2012).

The process of cortical folding, or gyrification, results in gyri and sulci that give the cortex its wrinkly appearance. The degree of cortical folding can be quantified using the local gyrification index (LGI) (Schaer et al., 2008). The LGI is a ratio of the total cortical surface to the superficially exposed outer surface tightly wrapping the cortex without entering the sulci. Cortical folding is a uniquely mammalian solution to increasing cortical grey matter without exaggerating head size. This process is also key to the optimization of axonal wiring and the functional organization of the brain (Klyachko & Stevens, 2003). The mechanisms of cortical folding are under active investigation. Recent perspectives suggest that tightly coordinated molecular genetic processes (Borrell, 2018) and biomechanical forces (Kroenke & Bayly, 2018) are involved. Radial expansion of progenitor cells might be particularly significant (see Fernández, Llinares-Benadero, & Borrell, 2016 for a review). Importantly, cortical folding provides a window on early development (Schaer et al., 2009). The major folding patterns are determined largely before birth and finish undergoing the most rapid morphological changes by childhood (Armstrong, Schleicher, Omran, Curtis, & Zilles, 1995 Jan-Feb). This sensitive period of neurodevelopment overlaps with the timing of the most prominent environmental risk factors associated with psychosis (Zwicker, Denovan-Wright, & Uher, 2018).

Large multisite neuroimaging studies have found reductions in cortical folding

among adults with psychotic disorders (Nanda et al., 2014; Nesvåg et al., 2014). Aberrant gyrification has also been reported in people at genetic risk for schizophrenia (Nanda et al., 2014; Falkai et al., 2007). By extension, recent work has shown that scores indexing genetic liability to schizophrenia are associated with regional reductions in the LGI (Liu et al., 2017). The prefrontal cortex has long been implicated in schizophrenia (Weinberger, 1988). Early work that measured cortical folding with manual tracing of MRI slices has shown lower frontal cortical folding in patients with schizophrenia (Kulynych, Luevano, Jones, & Weinberger, 1997). Correspondingly, recent work using automated 3D methods to quantify LGI have shown similar reductions in prefrontal cortical folding (Palaniyappan, Mallikarjun, Joseph, White, & Liddle, 2011).

The reported prefrontal LGI abnormalities are consistent with neuropathological findings from other imaging modalities and postmortem data (Kelly et al., 2017; Selemon, Kleinman, Herman, & Goldman-Rakic, 2002). Abnormal gyrification also predicts poor treatment response in first-episode psychosis (Palaniyappan et al., 2013) and has been used to distinguish patients with more severe illness from those with milder forms (Guo et al., 2015). Taken together, the body of literature suggests cortical folding alterations across the psychosis spectrum. However, whether or not these alterations are present before illness onset has been more difficult to establish.

Adolescence has been described as a critical period of vulnerability for schizophrenia (Selemon & Zecevic, 2015). However, the majority of clinical high-risk studies have focused on adulthood (Fusar-Poli et al., 2012). As such, the etiology of psychosis can be clarified by examining cortical folding earlier, before the onset of a functionally impairing illness. In the current study, we addressed this gap by examining the LGI in adolescents from a cohort with enriched familial risk who had experienced psychotic symptoms but did not meet the criteria for psychotic illness. We hypothesized that psychotic symptoms would be related to lower cortical folding in symptomatic youth, particularly in the prefrontal cortex.

3.3 Methods

3.3.1 Participants

As part of the ongoing Families Overcoming Risks and Building Opportunities for Well-being (FORBOW) study, we collected MRI scans from 110 participants aged 9–25 years. The FORBOW study is a longitudinal study enriched for the offspring of parents with mental illness (Uher et al., 2014). Those at familial risk for mental illness and participants from control families were invited to complete the MRI study. The study protocol was approved by the research ethics board of the Nova Scotia Health Authority. Participants provided written informed consent. For children who did not have the capacity to make a fully informed decision, a parent or guardian provided written informed consent and the child provided assent. Exclusion criteria were a personal history of psychotic illness, any serious medical or neurologic disorder, substance abuse or dependence during the previous 6 months, or MRI contraindications.

3.3.2 Participant clinical and cognitive assessment

Parent assessment

We used the Schedule for Affective Disorders and Schizophrenia (SADS-IV) (Endicott & Spitzer, 1978) and the Structured Clinical Interview for DSM-5 (SCID-5) (First, 2015) to establish diagnoses of mental disorders and psychosis according to DSM-IV and DSM-5. Diagnoses were confirmed in consensus meetings with a psychiatrist blind to offspring psychopathology.

Offspring assessment

Participating youth were interviewed using the Kiddie Schedule for Affective Disorders and Schizophrenia, Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997). Offspring assessors were blinded to parent psychopathology. Full-scale intelligence quotient (FSIQ) was assessed using the Wechsler Abbreviated Scale of Intelligence, second edition (Wechsler, 2011). Psychotic symptoms were assessed using the following instruments: the K-SADS-PL interview psychosis module and appendix,

consensus-rated by child and adolescent psychiatrists blind to parent psychopathology; in participants aged 3–12 years, the Structured Interview for Prodromal Syndromes (SIPS) (Miller et al., 2003), measuring attenuated psychotic symptoms; the "Funny Feelings" interview (Arseneault et al., 2011; Poulton et al., 2000; Polanczyk et al., 2010) (we included only youth with symptoms rated as "definite psychotic symptom" by consensus between 2 independent raters); the Schizophrenia Proneness Instrument, Child and Youth version (SPI-CY) (Fux, Walger, Schimmelmann, & Schultze-Lutter, 2013) for those aged 8 years and older, to assess abnormal experiences in the domains of perception, cognition, language and affect that strongly predict the development of psychosis (Klosterkötter et al., 2001; Schultze-Lutter et al., 2012).

Consistent with our previous reports (MacKenzie et al., 2016, 2017), we created a dichotomous variable for the presence of psychotic symptoms captured by any one of the following: confirmed hallucinations or delusions on K-SADS-PL, positive symptoms on SIPS rated ≥ 3, definite psychotic symptoms confirmed through independent curation, and high-risk basic symptom profiles of cognitive/perceptive disturbances on the SPI-CY.

3.3.3 MRI acquisition

Images were acquired with a 3 T General Electric Discovery MR750 scanner equipped with a 32-channel MR Instruments radiofrequency head coil. Scanning took place at the Biomedical Translational Imaging Centre (BIOTIC) in Halifax, Nova Scotia, Canada. Each participant was positioned supine in the MRI scanner, with their head supported by foam padding to reduce movement. Ear plugs were provided to minimize scanner noise. We collected a 3D T_1-weighted brain volume imaging (BRAVO) sequence with whole-brain coverage; 1 mm^3 isotropic resolution; matrix 224×224; field of view 224 mm; 168 sagittal slices at 1 mm thickness; repetition time 5.9 ms; echo time 2.2 ms; inversion time 450 ms; flip angle 12°; receiver bandwidth $\pm$ 62.5; number of excitations = 2; autocalibrating reconstruction for cartesian imaging (ARC) phase acceleration = 2; ARC slice acceleration = 1; no phase wrap; scan duration 5 minutes, 42 seconds.

3.3.4 MRI analysis

Data were preprocessed using the open-source Human Connectome Pipeline (Glasser et al., 2013). As part of the validated pipeline, we reconstructed the T_1-weighted scan surface using FreeSurfer version 5.3 (Fischl, 2012). An automated labelling system subdivided the cerebral cortex into gyral-based parcellations corresponding to the Desikan–Killiany atlas (Desikan et al., 2006).

Local gyrification index We quantified the 3D LGI from FreeSurfer output, as the ratio of the total cortical surface area (including cortex buried in sulci) to the outer cortex surface area, which tightly wraps the brain but does not enter the sulci. Thus, a higher LGI represents more cortical folding in each brain parcel, and a lower LGI represents less cortical folding. Details of the automated LGI computation can be found in the validation paper (Schaer et al., 2008) and on the FreeSurfer website

3.3.5 Statistical analysis

We performed statistical analyses in R Studio (version 3.5.0) (Team & Others, 2015). We compared demographic and clinical variables using t tests for continuous variables and χ^2 tests for categorical variables. The LGI for each of the 34 cortical parcellations across both hemispheres served as the primary dependent variable. The primary independent variable was the presence or absence of psychotic symptoms.

We tested the relationship between lifetime psychotic symptoms and LGI using mixed-effect generalized linear models. We accounted for the non-independence of brain data from related individuals by including the family identifier as a random effect. We included age and sex as covariates in the model. We controlled for multiple comparisons across brain parcellations using false discovery rate (FDR) (Benjamini & Hochberg, 1995). We reported effect sizes using standardized regression estimates (β) and their 95% confidence intervals (CIs).

To ensure that the observed relationship between cortical folding and psychotic symptoms was not due to the use of psychoactive substances or other factors linked to changes in brain structure, we conducted sensitivity analyses. In the brain regions found to be significant after correcting for multiple testing, we further covaried for lifetime cannabis use, lifetime stimulant use, FSIQ and estimated total intracranial

Table 3.1: Demographic and clinical characteristics of the study sample

Characteristic	No psychotic symptoms	Psychotic symptoms	*p* value
Participants, no.	62	48	—
Female sex, no. (%)	31 (50.0)	33 (68.8)	0.08
Mean age ± SD, yr	13.7 ± 3.4	14.5 ± 4.1	0.28
Siblings, no. (%)	11 (17.8)	12 (25.0)	0.49
Anxiety disorder, no. (%)	21 (33.9)	25 (52.1)	0.09
Parent diagnosis, no. (%)			0.12
None	18 (29.0)	6 (12.5)	—
Depression	23 (37.1)	27 (56.3)	—
Bipolar disorder	17 (27.4)	13 (27.1)	—
Schizophrenia	4 (6.5)	2 (4.2)	—
Parent psychosis, no. (%)	11 (17.8)	9 (18.8)	1.00
Cannabis use, no. (%)	7 (11.3)	6 (12.5)	1.00
Stimulant use, no. (%)	2 (3.2)	6 (12.5)	0.08
Mean FSIQ score ± SD	106.2 ± 14.3	107.9 ± 11.3	0.51
Mean eTIV ± SD, cm^3	1522.2 ± 138.6	1427.8 ± 168.0	0.002

eTIV = estimated total intracranial volume; FSIQ = Full Scale Intelligence Quotient; SD = standard deviation.
*We compared demographic and clinical variables using *t* tests for continuous variables and χ^2 tests for categorical variables.

volume (eTIV).

3.4 Results

3.4.1 Demographic variables

Of the 110 youth scanned, 48 (43.64%) met the criteria for a definite psychotic symptom on 1 or more assessments. Table 3.1 summarizes demographic and clinical characteristics by symptom status. General cognitive ability did not differ between participants with or without psychotic symptoms (t = 0.67, p = 0.51). We found a statistically significant difference in eTIV (t = -3.15, p = 0.002): youth with psychotic symptoms showed smaller eTIV (mean 1427.8 ± 168.0 cm^3) than youth without symptoms (1522.2 ± 138.6 cm^3).

Figure 3.1: Differences in mean cortical folding across the anatomical boundaries of the Desikan–Killiany atlas. Dark blue dots represent youth with psychotic symptoms (n = 48), light blue dots represent those without (n = 62). We found an overall trend toward lower cortical folding in youth with psychotic symptoms. (A) Regions of interest with mean local gyrification index < 3.0. (B) Regions of interest with mean local gyrification index > 3.0.

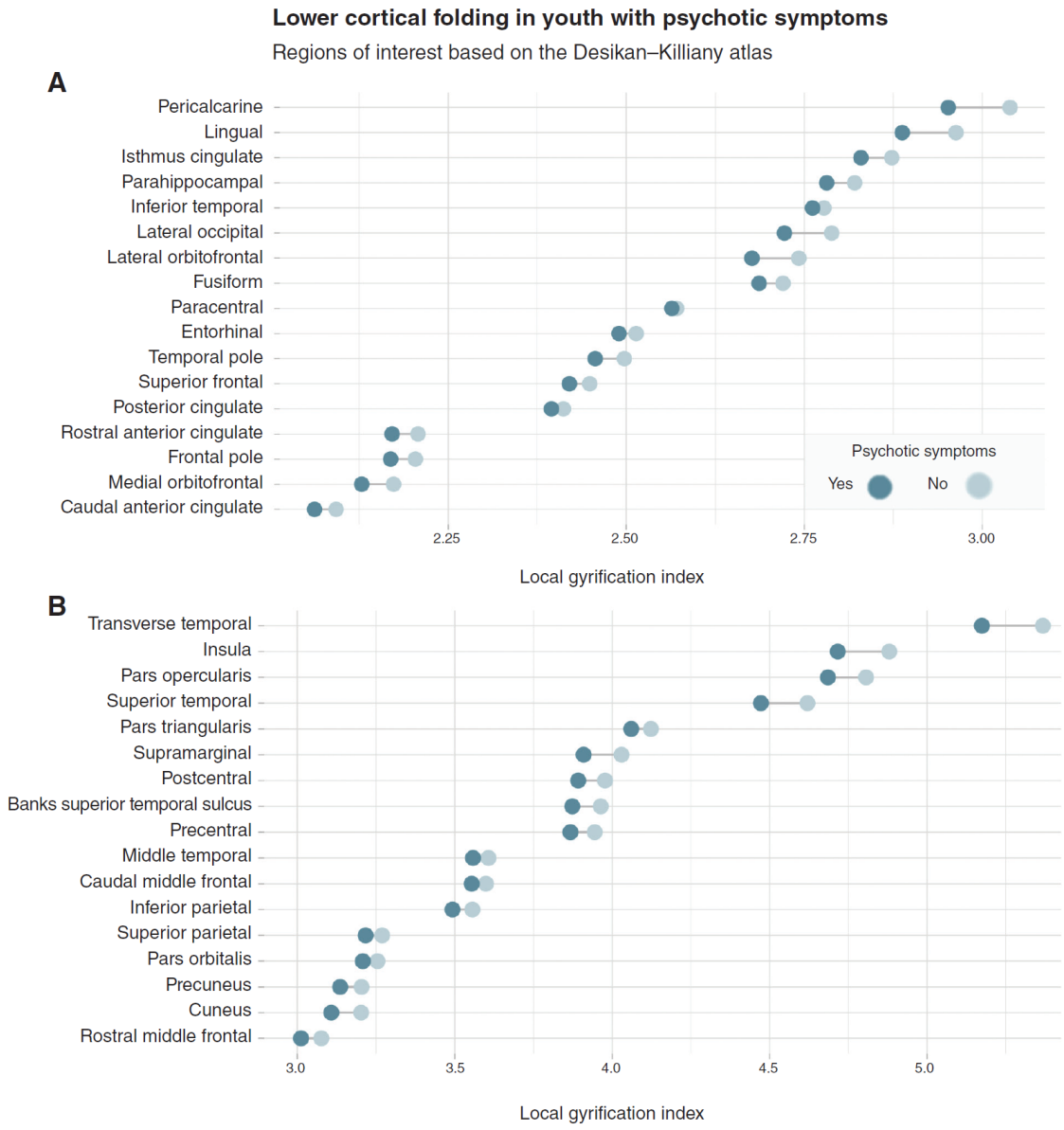
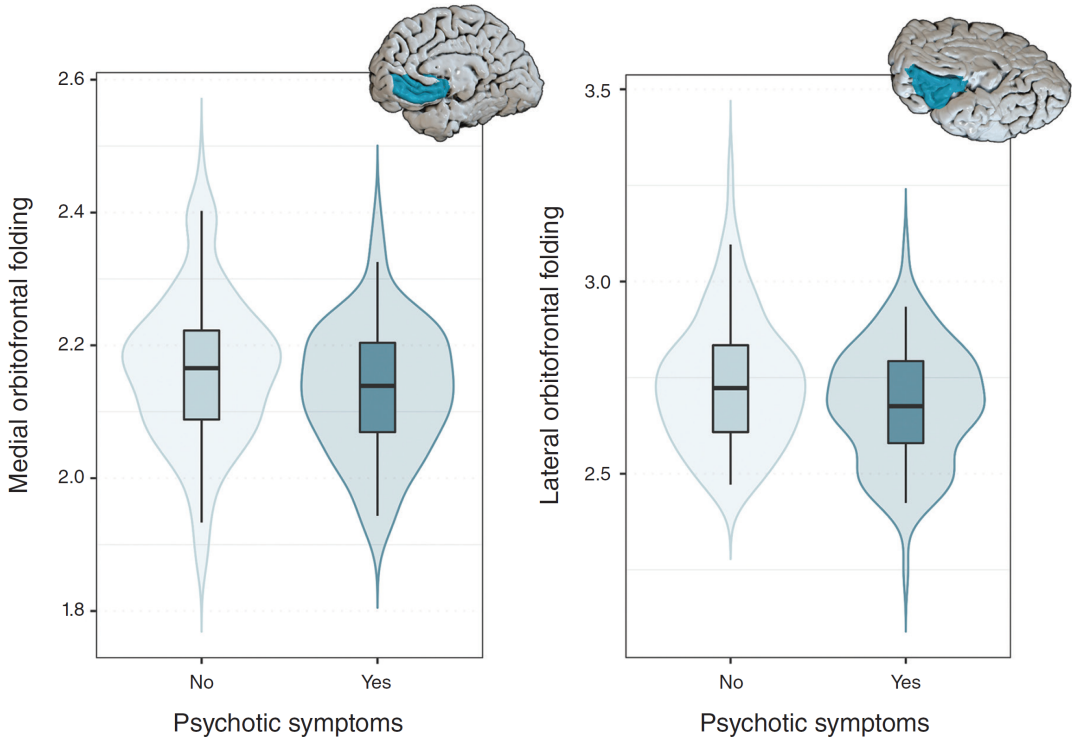


Figure 3.2: Lower prefrontal cortical folding in youth with psychotic symptoms (n = 48) versus those without (n = 62). Symptomatic youth in darker blue. Violin plots present cortical folding distributions by symptom status. Box plots nested within display median differences.



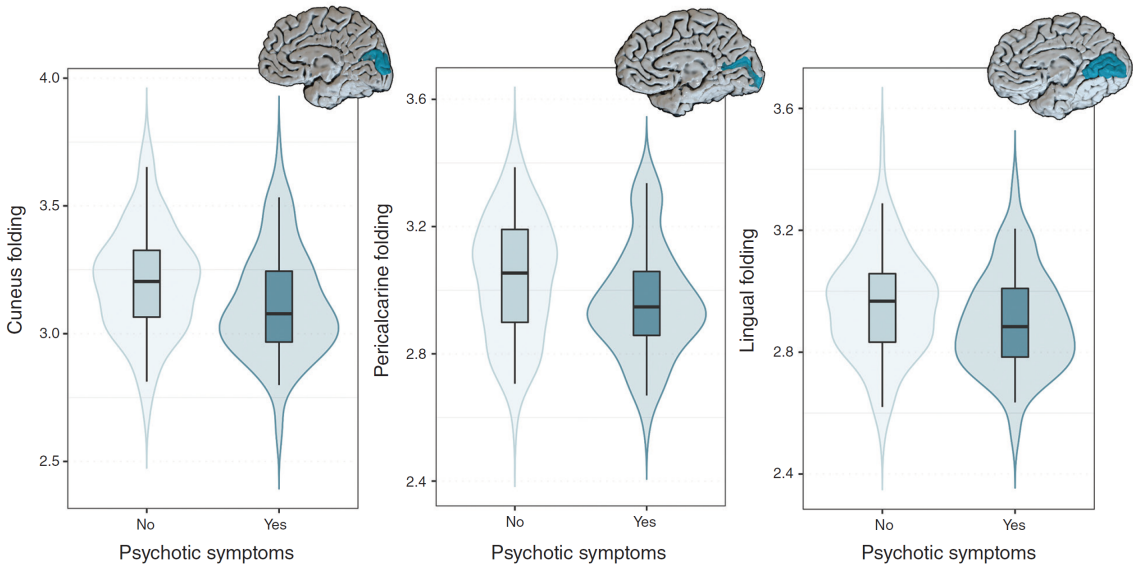
3.4.2 Overall differences in cortical folding across brain regions

We examined overall differences in cortical folding averaged across all brain regions. In a model controlling for sex, age and familial clustering, average whole-brain cortical folding was lower in youth with psychotic symptoms (β = -0.13, 95% CI -0.21 to -0.05, $p = 0.001$). This effect was no longer statistically significant following covariance for eTIV (β = -0.05, 95% CI -0.13 to 0.03, p = 0.22). Figure 3.1 shows the mean differences in folding across all structures.

3.4.3 Prefrontal cortical folding

We used the same model to explore regional differences in cortical folding. In line with our hypothesis, the exploratory analysis revealed lower prefrontal cortical folding in

Figure 3.3: Lower occipital cortical folding in youth with psychotic symptoms (n = 48) versus those without (n = 62). Symptomatic youth in darker blue. Violin plots present cortical folding distributions by symptom status. Box plots nested within display median differences.



youth with psychotic symptoms (Figure 3.2). Specifically, the mixed effect generalized linear models controlling for sex, age and familial clustering localized the lower cortical folding to the medial (β = -0.42, 95% CI -0.71 to -0.12, $p = 0.006$, $p_{FDR} = 0.04$) and lateral (β = -0.39, 95% CI -0.66 to -0.12, $p = 0.005$, $p_{FDR} = 0.04$) aspects of the orbitofrontal cortex (OFC).

In sensitivity analyses of this finding, youth with psychotic symptoms robustly showed lower OFC folding in models controlling for cannabis use, stimulant use, FSIQ and eTIV, both medially (β = -0.34, 95% CI -0.63 to -0.04, $p = 0.025$) and laterally (β = -0.28, 95% CI -0.54 to -0.01, $p = 0.042$).

3.4.4 Occipital lobe cortical folding

Along with the OFC findings, the exploratory analysis revealed three additional regions that survived brain-wide correction for multiple comparisons in models controlling for age, sex and familial clustering (Figure 3.3). Psychotic symptoms were associated with lower regional gyrification in the occipital lobe, specifically the cuneus

(β = -0.47, 95% CI -0.75 to -0.19, $p = 0.001$, $p_{FDR} = 0.03$), the pericalcarine gyrus (β = -0.45, 95% CI -0.73 to -0.17, $p = 0.002$, $p_{FDR} = 0.03$) and the lingual gyrus (β = -0.38, 95% CI -0.66 to -0.11, $p = 0.006$, $p_{FDR} = 0.04$). We saw the same pattern of results when examining symptoms from a dimensional perspective (Supplement Figure A.2).

We conducted sensitivity analyses to determine if the differences in cuneus, pericalcarine and lingual cortical folding might be attributable to extraneous variables rather than to symptom status. Again, we implemented models identical to the exploratory analysis while simultaneously covarying for cannabis use, stimulant use, FSIQ and eTIV. As with the prefrontal results, the link between psychotic symptoms and occipital folding remained significant for the cuneus (β = -0.46, 95% CI -0.74 to -0.18, $p = 0.001$) and the pericalcarine (β = -0.40, 95% CI -0.69 to -0.11, $p = 0.007$) and lingual (β = -0.35, 95% CI -0.63 to -0.07, $p = 0.015$) gyri.

3.5 Discussion

We sought to determine whether youth with psychotic symptoms displayed cortical aberrations before the onset of impairing psychotic illness. To answer that question, we examined 3D reconstructions of cortical folding, an early neurodevelopmental marker of cortical expansion. We found a pattern of lower cortical folding in adolescents who had psychotic symptoms but who did not meet the criteria for a psychotic disorder.

3.5.1 Lower cortical folding

In our study, psychotic symptoms were related to lower cortical folding across all brain regions, with statistically significant regional effects. This unidirectional pattern is supported by the literature on brain structure among adult patients with schizophrenia-spectrum disorders. Nesvåg et al., 2014 examined cortical folding among 207 patients with schizophrenia and 206 controls. They found that patients had a lower LGI in large clusters of the cerebral cortex, leading them to conclude that reduced gyrification is a feature of the brain pathology in schizophrenia. Similar to our work, no regions had significantly higher LGI among patients. Nanda et al., 2014 examined the LGI in 388 patients with psychotic disorders and 243 controls. Their

multicentre study found that patients with psychotic disorders had a significantly lower LGI than controls. Importantly, the directionality of the finding was consistent with our findings, in that no regions were less folded among controls.

3.5.2 Differences in prefrontal cortex folding

Lower prefrontal cortical folding was localized to the medial and lateral OFC. Our finding was in line with previous work implicating abnormalities in the OFC cortical folding pattern in psychosis (Palaniyappan et al., 2013; Bartholomeusz et al., 2013), first-episode schizophrenia (Takayanagi et al., 2010) and chronic schizophrenia (Isomura et al., 2017). In our adolescent sample, this difference was unlikely to be attributable to common confounders in psychosis research, such as illness burden or the effects of medication. The localization to the OFC in our study was consistent with prospective studies in patients at clinical high risk and in animal models. Patients at clinical high risk who convert to psychosis show steeper rates of cortical thickness decline and grey matter reduction in the OFC than non-converters at clinical high risk and controls (Pantelis et al., 2003; Cannon et al., 2015). Furthermore, OFC neurons have been shown to be common targets for both typical and atypical antipsychotic drugs (Homayoun & Moghaddam, 2008).

Characteristics of schizophrenia include aberrant perception and cognitive deficits. The OFC is involved in a number of disorder-related functions, such as sensory integration, learning, and social and emotional decision-making (Rolls & Grabenhorst, 2008; Seo & Lee, 2012). Our previous work examining cognition showed that youth with psychotic symptoms exhibited deficits in executive functioning (MacKenzie et al., 2017). We specifically found impaired emotional decision making, even after controlling for general cognitive ability. The combination of a reduction in OFC folding and impaired decision-making may be a neurocognitive marker of a propensity for psychotic symptoms.

3.5.3 Differences in occipital cortex folding

We found lower cortical folding in the cuneus and the pericalcarine and lingual gyri. The cuneus is located between the calcarine and parieto-occipital sulci. The pericalcarine gyrus can be visualized between the cuneus and lingual gyrus on a midline

view of the brain. The lingual gyrus sits within the tentorial surface of the occipital lobe, inferior to the calcarine sulcus.

Studies of cortical folding and functional connectivity in psychosis have found reduced cortical folding in the lingual cortex of patients with psychosis compared with healthy controls (Palaniyappan et al., 2013; Palaniyappan & Liddle, 2014). They also showed that the aberrant functional connectivity of the visual processing regions was a better predictor of symptom persistence and burden than diagnostic information. Our current work and aforementioned findings also correspond to studies examining volumetric data. Compared to healthy controls, patients with first-episode schizophrenia demonstrated significantly reduced grey matter volumes in the lingual gyrus (Chang et al., 2016). Significant cortical thinning has also been noted in this region in patients diagnosed with schizophrenia compared to matched controls (Schultz et al., 2013; Kong et al., 2015). Finally, 22q11 deletion syndrome and genetic risk for schizophrenia based on common genetic variation have been associated with volumetric grey matter differences in the lingual gyrus and cuneus, as well as cortical thinning in the cuneus and pericalcarine and lingual gyri (C. A. Thompson et al., 2017; Walton et al., 2018).

A recent paper examined LGI in medicated young adults who met the criteria for an at-risk mental state (Sasabayashi et al., 2017). In contrast to our findings of decreased LGI, those in an at-risk mental state showed widespread increases in LGI. Interestingly, increased gyrification in the cuneus, pericalcarine and lingual regions was related to risk for transition to a psychotic disorder. This work underscores the relevance of occipital cortical folding in the risk for psychosis, but the discrepancy in the directionality of findings indicates the need for further longitudinal study of the developmental transition period from adolescence to young adulthood in high-risk populations.

3.5.4 Limitations

This study had several limitations. The onset of schizophrenia and other psychotic disorders typically occurs in late adolescence or early adulthood. In our developmental sample, the age range overlapped with this period. In other words, our sample

included participants younger than the age with the highest risk of onset, and participants passing through this stage. Future studies can reduce this heterogeneity, particularly in cohorts examining brain development at fixed age ranges. Furthermore, the cross-sectional nature of this work limited our ability to track prospective developmental changes in regions with reduced folding. To address some of these challenges, we propose the collection of longitudinal imaging follow-up data.

We were able to minimize confounds of illness burden, comorbidities and medication use by studying youth. Nevertheless, a proportion of the sample had been exposed to marijuana and psychoactive medication, both of which could have an effect on the developing brain (Shollenbarger, Price, Wieser, & Lisdahl, 2015; Spencer et al., 2013). Sensitivity analyses suggested that these substances did not affect the association between reduced cortical folding and psychotic symptoms. Cannabis use remains an important covariate to control for, because adolescent initiation of cannabis use has been linked to early-onset psychosis (Bagot, Milin, & Kaminer, 2015), and legalization of the substance may affect initiation or usage.

We found a significant difference in total intracranial volume in our sample. Youth with psychotic symptoms had lower eTIV. Although this finding was not part of our hypothesis, there is meta-analytic evidence for reduced intracranial and total brain volume in schizophrenia (Haijma et al., 2013). Because certain structures scale with intracranial volume (Buckner et al., 2004), we controlled for eTIV as a covariate. This correction eliminated the overall average difference in cortical folding across the brain, but the regional differences in frontal and occipital cortical folding remained robust to this correction. Future work is needed to contextualize the clinical relevance of eTIV differences in samples of risk-enriched youth.

Finally, we examined a single neurodevelopmental marker, but we know that cortical folding is related to optimization of axonal wiring and functional organization in the brain (Klyachko & Stevens, 2003). Future research should integrate additional imaging data, such as probabilistic tractography, intracortical myelination and resting-state functional connectivity. For example, one multi-analysis study showed that decreased frontal gyrification in adolescent schizophrenia may be associated with widening of the frontal sulci and reductions in cortical surface area (Janssen et al., 2014). Multimodal extension and synthesis with molecular genetic and neurocognitive

data will bring new insights into the significance of our findings.

3.6 Conclusion

This study found regional reductions in cortical folding of adolescents who had experienced psychotic symptoms. The young age of the cohort helped to clarify the neurodevelopmental basis of psychosis at an early stage, before medication, drug use and illness burden could take a persistent toll on the brain.

Chapter 4

Reliability of multimodal MRI brain measures in youth at risk for mental illness

Copyright statement This chapter is based on a manuscript that has been published: Vladislav Drobinin, Holly Van Gestel, Cark A Helmick, Matthias H Schmidt, Chris V Bowen, and Rudolf Uher. Reliability of multimodal MRI brain measures in youth at risk for mental illness (2020). *Brain and Behavior*. Re-use is permitted with copyright permission (Appendix B.3).

Contribution statement I drafted the manuscript used as the basis for this chapter, with guidance and editing from Dr. Rudolf Uher, and the other co-authors. Data were collected by Holly Van Gestel with the help of myself and other FORBOW staff. I completed the data analysis.

4.1 Abstract

Introduction A new generation of large-scale studies is using neuroimaging to investigate adolescent brain development across health and disease. However, imaging artifacts such as head motion remain a challenge and may be exacerbated in pediatric clinical samples. In this study, we assessed the scan–rescan reliability of multimodal MRI in a sample of youth enriched for risk of mental illness.

Methods We obtained repeated MRI scans, an average of 2.7 ± 1.4 weeks apart, from 50 youth (mean age 14.7 years, SD = 4.4). Half of the sample (52%) had a diagnosis of an anxiety disorder; 22% had attention-deficit/hyperactivity disorder (ADHD). We quantified reliability with the test–retest intraclass correlation coefficient (ICC).

Results Gray matter measurements were highly reliable with mean ICCs as follows: cortical volume (ICC = 0.90), cortical surface area (ICC = 0.89), cortical thickness (ICC = 0.82), and local gyrification index (ICC = 0.85). White matter volume reliability was excellent (ICC = 0.98). Diffusion tensor imaging (DTI) components were also highly reliable. Fractional anisotropy was most consistently measured (ICC = 0.88), followed by radial diffusivity (ICC = 0.84), mean diffusivity (ICC = 0.81), and axial diffusivity (ICC = 0.78). We also observed regional variability in reconstruction, with some brain structures less reliably reconstructed than others.

Conclusions Overall, we showed that developmental MRI measures are highly reliable, even in youth at risk for mental illness and those already affected by anxiety and neurodevelopmental disorders. Yet, caution is warranted if patterns of results cluster within regions of lower reliability.

4.2 Introduction

A new generation of large-scale studies (Alexander et al., 2017; Casey et al., 2018) is using neuroimaging techniques to investigate adolescent brain development across health and disease. These tremendous undertakings, often called "biobanks" for the wealth of biological data that they collect, are particularly focused on mental health. A primary goal includes identifying developmental trajectories of psychiatric illness which in turn might help improve early detection and guide intervention (Alexander et al., 2017; Casey et al., 2018). Such research is highly valuable, as epidemiologic studies show that 75% of psychiatric disorders begin early in the lifespan, prior to age 24 (Kessler et al., 2005; Kim-Cohen et al., 2003). However, identifying clinically useful brain markers of illness, or "biomarkers," hinges on the reliability of the MRI data.

Reliability is the ability of a measurement to provide consistent results under similar circumstances. Imaging artifacts, such as head motion, remain a challenge to reliability (Reuter et al., 2015), and there are concerns that measurement error may be exacerbated in pediatric clinical samples (Ducharme et al., 2016). Functional MRI studies have begun to address within-subject reliability in youth as motion can have a profound effect on functional connectivity estimates (Van Dijk, Sabuncu, & Buckner, 2012; Vetter et al., 2017). However, a large body of imaging research deals with brain structure, and here too image artifacts are of concern. It has been shown that head motion in healthy volunteers can resemble cortical gray matter atrophy (Reuter et al., 2015). Children and adolescents might be particularly sensitive to scanner noise and may have difficulty remaining still for the duration of the sequences. One study examining pediatric MRI data has shown that low-quality data can affect inferences regarding the developmental trajectories of cortical maturation (Ducharme et al., 2016). These findings necessitate the assessment of the reliability of MRI data in participants who are not merely undergoing normal development but are also showing externalizing and internalizing symptoms or are at increased familial risk for mental illness (Rasic et al., 2014).

In this study, we assessed the scan–rescan reliability of multimodal MRI in a

sample of youth at risk for mental illness, including those already experiencing psychopathology. We measured common structural imaging metrics reported in the literature and quantified regional reliability based on widely used brain atlases. We also compared the reliability of structural measures to published estimates from samples of healthy adults.

4.3 Materials and methods

4.3.1 Participants

We recruited 53 youth (mean age 14.7 years, SD = 4.4) at familial risk for mental illness from FORBOW study, a longitudinal study enriched for sons and daughters of parents with mental illness (Uher et al., 2014). Offspring at familial risk for mental illness and participants from control families were invited to complete the MRI study. Participants were scanned twice, an average of 2.7 ± 1.4 weeks apart. Exclusion criteria were personal history of (i) psychotic illness, (ii) any serious medical or neurologic disorders, or (iii) MRI contraindications. The study protocol was approved by the Research Ethics Board of the Nova Scotia Health Authority. Participants provided written informed consent. For children who did not have capacity to make a fully informed decision, a parent or guardian provided written informed consent and the child provided assent.

4.3.2 Parent assessment

We used the Schedule for Affective Disorders and Schizophrenia (SADS-IV; Endicott & Spitzer, 1978) and the Structured Clinical Interview for DSM-5 (SCID-5; First, 2015) to establish diagnoses of mental disorders according to DSM-IV and DSM-5.

4.3.3 Offspring assessment

Participating youth were interviewed using the Kiddie Schedule for Affective Disorders and Schizophrenia, Present and Lifetime Version (K-SADS-PL; Kaufman et al., 1997) by assessors blind to parent psychopathology. Diagnoses were confirmed in consensus meetings with a psychiatrist. Full-scale intelligence quotient (FSIQ) was assessed using the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011).

4.3.4 Socioeconomic status (SES)

Socioeconomic status was captured as a composite variable (range 0-5) indexing: (i) maternal and (ii) paternal levels of education (iii) family household annual income, (iv) ownership of primary residence, and (v) ratio of bedrooms to residents in household, as previously described (MacKenzie et al., 2017; Zwicker, MacKenzie, et al., 2019). Higher numeric value reflects higher SES.

4.3.5 MRI acquisition

Images were acquired with a 3T General Electric Discovery MR750 scanner equipped with a 32-channel MR Instruments RF head coil. Scanning took place at the Biomedical Translational Imaging Centre (BIOTIC), Halifax, Nova Scotia. Each participant was positioned supine in the MRI scanner with the head supported by foam padding to reduce movement. Earplugs were provided to minimize scanner noise. We collected a 3D T_1-weighted (T1w) Brain Volume imaging (BRAVO) sequence with whole-brain coverage, 1 mm^3 isotropic resolution, matrix = 224 × 224, field of view (FOV) = 224 mm, 168 sagittal slices at 1 mm thickness, repetition time (TR) = 5.9 ms, echo time (TE) = 2.2 ms, inversion time (TI) = 450 ms, flip angle = 12°, receiver bandwidth = ±62.5, number of excitations (NEX) = 2, autocalibrating reconstruction for cartesian imaging (ARC) phase acceleration = 2, ARC slice acceleration = 1, no phase wrap, scan duration = 5 min 42 s.

In addition, we collected a T_2-weighted fluid attenuated inversion recovery (FLAIR) sequence using a T2 prep contrast option (T2PREP) with identical coverage, resolution and acquisition orientation to the T1w sequence, TE = 98 ms, TR = 5,100 ms, TI = 1,427 ms, echo train length (ETL) = 250 echoes, flip angle = 90°, receiver bandwidth = ±62.5 kHz, NEX = 1, with prospective motion correction (PROMO) enabled (White, Su, Schmidt, Kao, & Sapiro, 2010), ARC phase = 2.5, ARC slice = 1, scan duration = 5 min.

Whole-brain axial–oblique diffusion-weighted images were also acquired using a single-shot spin-echo EPI pulse sequence, gradient directions = 30, b-value = 1,000 s/mm^2, three b = 0 images interleaved every 15 volumes, TR = 8,000 ms, TE = 66.7, FOV = 216 mm, slice thickness = 2 mm, number of slices 76, matrix = 108 ×

108, voxel = 23 mm isotropic, receiver bandwidth = ±250 kHz, ASSET phase acceleration factor = 2, phase-encode direction = AP, scan duration = 4 min 32 s. For the purposes of estimating and correcting susceptibility-induced distortions, we also acquired a second whole-brain axial–oblique diffusion-weighted sequence with matching parameters except only 8 volumes at b = 0 and with opposite phase-encoding direction = PA, scan duration = 64 s.

4.3.6 MRI processing

Scans were processed with the Human Connectome Project (HCP) Minimal Preprocessing Pipeline (Glasser et al., 2013). The HCP pipeline is a well-documented set of scripts developed to analyze high-quality multimodal MRI data. It leverages the most widely used open-source MRI processing software: FreeSurfer 6 (RRID:SCR_001847; Fischl, 2012) and the FMRIB Software Library (FSL, RRID:SCR_002823; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012).

We have optimized the pipeline for our data by matching it to our acquisition parameters and by replacing the MNI template with a pediatric template for registration. The modified pipeline is available and freely accessible on GitHub. We used the NIHPD pediatric atlas (NIHPD Objective 1 atlases [4.5–18.5 years], RRID:SCR_008794; Fonov et al., 2011) to minimize registration bias in our developmental cohort. In order to measure cortical folding, we ran the local gyrification index (LGI) analysis, the details of which can be found in the validation paper (Schaer et al., 2008) and on the FreeSurfer website.

For gray matter reliability, we examined 34 cortical regions of interest per hemisphere based on the FreeSurfer default Desikan–Killiany atlas (Desikan et al., 2006). Thus, we measured cortical gray matter volume, cortical surface area, cortical thickness, and LGI/cortical folding in 68 parcellations per individual at each time point. Quality control was done both manually early in the processing stream and later with an automated supervised-learning tool on the FreeSurfer segmented output. Manual quality ratings of T1-weighted and T2-weighted images were performed by authors VD and HVG. Automated quality control was done with the Qoala-T tool (Klapwijk, van de Kamp, van der Meulen, Peters, & Wierenga, 2019). Qoala-T is an automated machine learning model used to classify the quality of FreeSurfer output. Six scans

from three participants were excluded after the combined quality control (largely due to excess motion), bringing the total number of participants in the analysis to 50.

For white matter reliability, we examined white matter volume and diffusion tensor imaging (DTI) metrics based on the 20-structure JHU DTI-based white-matter tractography atlas (Mori, Wakana, van Zijl, & Nagae-Poetscher, 2005). Data inclusion required absolute and relative motion to be under one and a half times the voxel size. Briefly, the processing was done in three steps: (i) creating binary maps of the 20 tracts in MNI152-space, (ii) registering each binary map into subject diffusion–space by combining and applying the nonlinear warps from MNI152 to NIHPD space, and NIHPD space to subject T_1-weighted space, and the rigid-body linear transform from subject T_1-weighted space to subject diffusion–space, (iii) using "fslstats" to report each metric; white matter volume, fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) for each tract.

4.3.7 Statistical analysis

We used RStudio (R Version 3.6.2; RStudio version 1.2.5033; RStudio Team, 2019) to calculate the intraclass correlation coefficient (ICC) for the processed scan–rescan datasets. Reliability is the ability of a measurement to provide consistent results under similar circumstances. Test–retest reliability assesses stability under repeated tests, quantifying the extent to which measurements can be replicated. The ICC indexes both correlation and agreement between measurements (Koo & Li, 2016) and is commonly used to quantify reliability. We wanted to capture the variation in measurements taken by MRI and introduced in postprocessing, on the same participant under the same conditions weeks apart. We used ICC (1,1) for calculating scan–rescan reliability implemented in the ICC package (Version 2.3.0; Wolak, Fairbairn, & Paulsen, 2012) which estimates the ICC and confidence intervals using the variance components form a one-way ANOVA. We examined averaged ICC and the regional (parcellated) ICC for all measures and classified reliability according to generally defined criteria (Cicchetti, 1994): poor (<0.40), fair (0.41-0.59), good (>0.59-0.74), and excellent (>0.74). The code, data and analysis, is available in a reproducible R notebook; `https://github.com/GitDro/YouthReliability`. We also repeated the analysis on a subsample of the participants scanned again twice, on average 14 months following

their initial pair of scans (see Tables S1–S10).

4.4 Results

4.4.1 Demographic and clinical characteristics

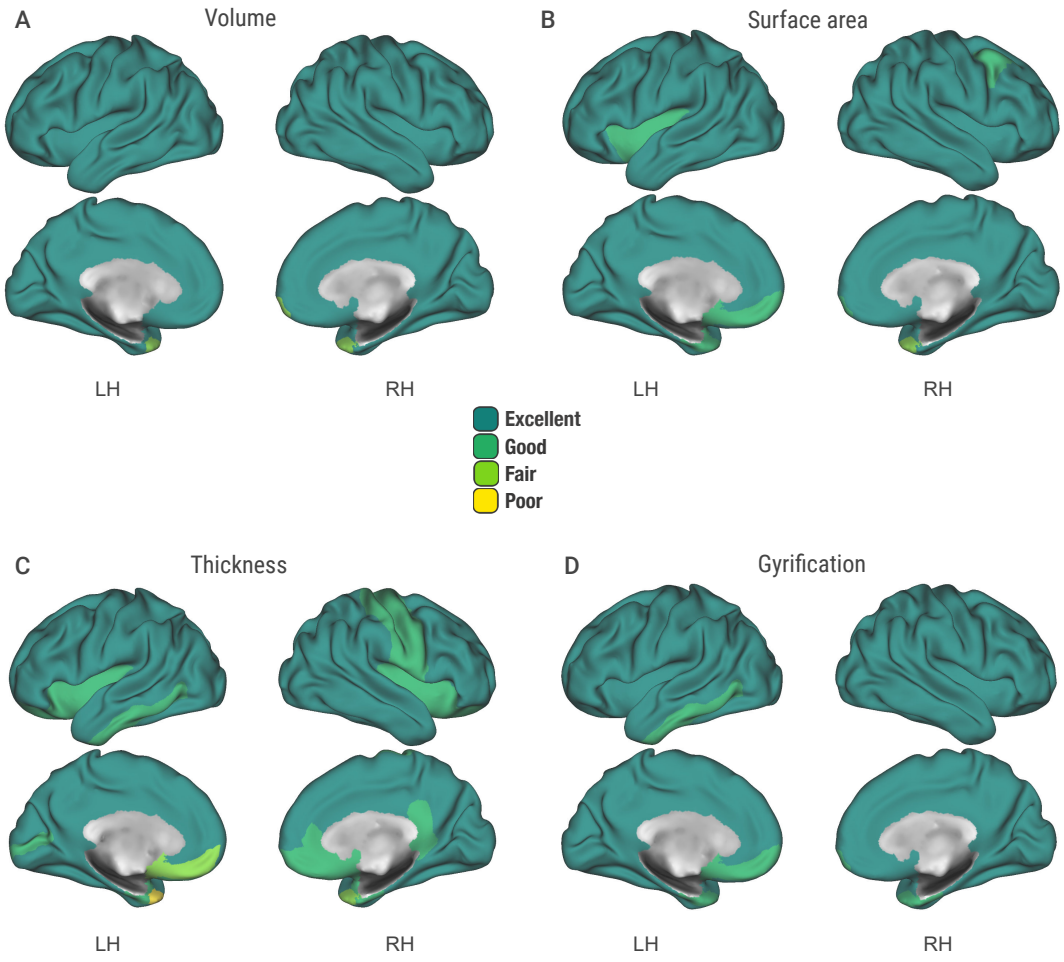
We present results from 100 scans collected from 50 youth (64% female) imaged several weeks apart (M = 2.70, SD = 1.36). The age range was 9–25 years old (M = 14.7, SD = 4.4). The majority of the participants have a family history of mental illness: 25 (50%) with a family history of major depressive disorder, 13 (26%) with a family history of bipolar disorder, and 2 (4%) with a family history of schizophrenia. Ten participants (20%) were recruited from control families. A large proportion of the scanned youth have been affected by mental illness: 26 participants (52%) had been diagnosed with an anxiety disorder, 13 (26%) had been diagnosed with major depressive disorder, and 11 (22%) have had a diagnosis of attention-deficit/hyperactivity disorder (ADHD).

The sample was predominantly white (90%), with a minority (10%) comprised of Indigenous and Black youth. The composite SES indicator was normally distributed (M = 3.1, SD = 1.31). Full-scale intelligence quotient (FSIQ) for the sample was in the normal range (M = 103, SD = 12.9).

4.4.2 Cortical volume

We observed "excellent" scan–rescan ICC for cortical gray matter volume (M = 0.90, 95% CI [0.84, 0.94]) averaged across the Desikan atlas regions (Figure 1a). The results were consistent across the left hemisphere (M = 0.92, 95% CI [0.86, 0.95]) and the right hemisphere (M = 0.89, 95% CI [0.82, 0.93]). As indicated by the high mean ICC, the reliability for most of the structures (65 out of 68; 96%) was classified as "excellent." However, there was some regional variation (Appendix Supplement Table A.1). The left supramarginal gyrus volume was the most reliably reconstructed, with near-perfect ICC (ICC = 0.99, 95% CI [0.98, 0.99]). The volume of the left temporal pole was the least reliably measured (ICC = 0.47, 95% CI [0.23, 0.66]), with the ICC dipping into the "fair" classification and the lower bound of the confidence interval crossing the "poor" threshold. The contralateral right temporal pole was the next

Figure 4.1: Reliability of cortical gray matter measures. Scan–rescan reliability of Desikan–Killiany regions. Intraclass correlation coefficient (ICC) values: poor (<0.40), fair (0.41–0.59), good (>0.59–0.74), and excellent (>0.74). (a) Cortical gray matter volume. (b) Cortical surface area. (c) Cortical thickness. (d) Local gyrification index



least reliably measured structure (ICC = 0.55, 95% CI [0.33, 0.72]). The only other structure with a designation below "excellent" was the right frontal pole, for which the ICC was only "fair" (ICC = 0.58, 95% CI [0.36, 0.74]). Cortical volume is a composite measure comprised of cortical surface area and cortical thickness; thus, we proceeded to examine the reliability of its components.

4.4.3 Cortical surface area

Averaged across the Desikan atlas regions, the ICC for cortical surface area (M = 0.89, 95% CI [0.82, 0.93]) was also deemed "excellent" overall (Figure 1b). Similar degree of reliability was attained both in the left hemisphere (M = 0.91, 95% CI [0.85, 0.95]) and the right hemisphere (M = 0.87, 95% CI [0.79, 0.92]). Just as with the volumetric results, the left supramarginal gyrus showed the highest ICC (ICC = 0.99, 95% CI [0.98, 0.99]). However, the ICCs for 12% of the Desikan regions were classified as "good" or "fair" (Appendix Supplement Table A.2). The bilateral temporal poles were the least reliably reconstructed structures (left; ICC = 0.65, 95% CI [0.45, 0.78], right; ICC = 0.47, 95% CI [0.23, 0.66]). The frontal poles also showed lower ICCs than most structures (left; ICC = 0.69, 95% CI [0.51, 0.81], right; ICC = 0.70, 95% CI [0.53, 0.82]). The left insula, entorhinal cortex, and medial orbitofrontal cortex were classified as "good" with respective ICCs of 0.72, 0.71, 0.64, 95% CI [0.55, 0.83], [0.55, 0.83], [0.45, 0.78]. Finally, the right caudal middle frontal gyrus ICC confidence interval ranged from "fair" to "excellent" (ICC = 0.70, 95% CI [0.53, 0.82]).

4.4.4 Cortical thickness

Across the Desikan atlas, the mean ICC for cortical thickness was "good" to "excellent" (M = 0.82, 95% CI [0.71, 0.89]). The results were consistent across the left hemisphere (M = 0.83, 95% CI [0.73, 0.90]) and the right hemisphere (M = 0.81, 95% CI [0.69, 0.88]). The regional variability was more apparent than for other measures (Figure 1c), with 24% of the atlas below the "excellent" reliability designation (Appendix Supplement Table A.3). Cortical thickness reconstruction was most reliable in the left superior frontal gyrus (ICC = 0.95, 95% CI [0.91, 0.97]). Once again, the temporal pole reconstruction was least reliable bilaterally (left; ICC = 0.38, 95% CI [0.12, 0.60], right; ICC = 0.41, 95% CI [0.16, 0.62]). Of note, two additional regions

had the lower bound of the confidence interval cross into "poor" reliability. This included the right entorhinal cortex (ICC = 0.60, 95% CI [0.39, 0.75]) and the left medial orbitofrontal cortex (ICC = 0.56, 95% CI [0.34, 0.73]).

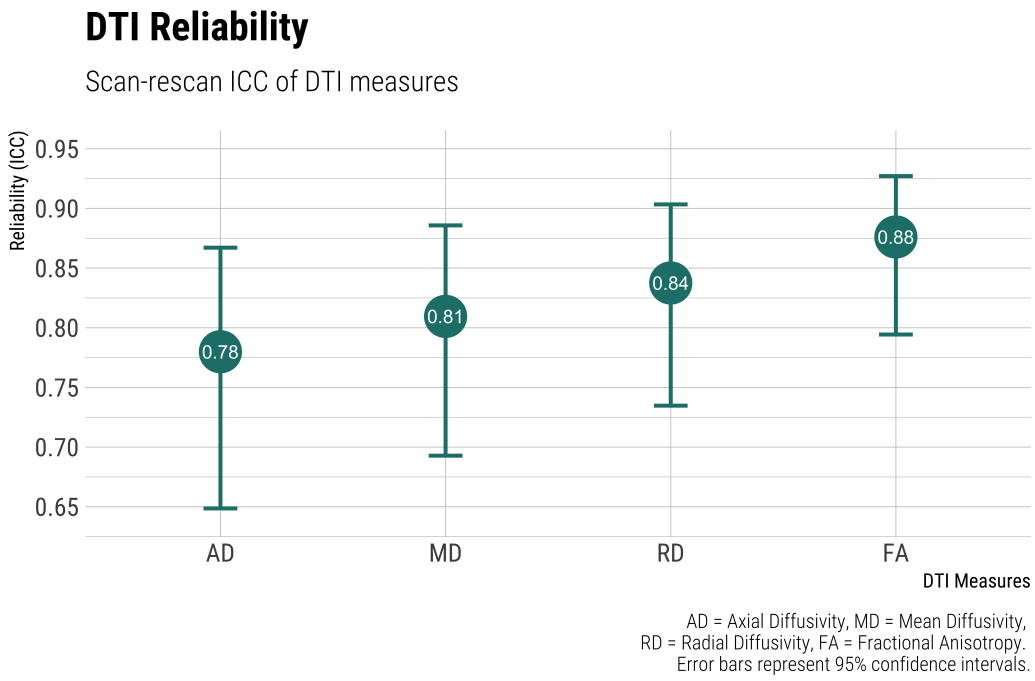
4.4.5 Cortical folding (LGI)

We also found "excellent" scan–rescan ICC (M = 0.85, 95% CI [0.75, 0.91]) for the measurement of cortical folding (Figure 1d). The average ICC across the right hemisphere was "excellent" (M = 0.85, 95% CI [0.76, 0.91]) and "good" to "excellent" across the left hemisphere (M = 0.84, 95% CI [0.74, 0.90]). Regional reliability was fairly consistent, with most structures displaying "excellent" reconstruction, and the rest (eight structures) achieving "good" ICCs. None of the confidence intervals dipped into the "poor" classification (Appendix Supplement Table A.4). Cortical folding reconstruction was most reliable in the right precentral gyrus (ICC = 0.95, 95% CI [0.92, 0.97]). As expected by now, the bilateral frontal (left; ICC = 0.62, 95% CI [0.41, 0.76], right; ICC = 0.66, 95% CI [0.48, 0.79]) and temporal poles (left; ICC = 0.63, 95% CI [0.43, 0.77], right; ICC = 0.68 95% CI [0.50, 0.81]) had the comparably lowest ICC estimates in measures of regional gyrification. The gyrification measurement of entorhinal cortex was also among the least reliable (left; ICC = 0.65, 95% CI [0.45, 0.78], right; ICC = 0.71, 95% CI [0.54, 0.82]), however, still "fair" to "excellent."

4.4.6 White matter volume

We observed remarkable reliability of white matter volume measurements averaged across the JHU white-matter tractography atlas (M = 0.98, 95% CI [0.97, 0.99]). Near-perfect reliability was observed in the reconstruction of the cingulum near the cingulate gyrus (ICC = 0.99, 95% CI [0.99, 1.00]) and other white matter tracts (Appendix Supplement Table A.5). The lowest regional reliability was observed in the cingulum near the hippocampus (ICC = 0.96, 95% CI [0.93, 0.98]), which was nevertheless categorized as "excellent."

Figure 4.2: Scan–rescan reliability of diffusion tensor imaging (DTI) measures.



4.4.7 Diffusion tensor imaging (DTI) measures

Fractional anisotropy (FA)

Next, we examined white matter FA, which is often used to index microstructural integrity. Scan–rescan ICCs were "excellent" averaged across the JHU atlas (M = 0.88, 95% CI [0.79, 0.93]), Figure 2. The left superior longitudinal fasciculus was measured most reliably across scan sessions (ICC = 0.95, 95% CI [0.91, 0.97]). The forceps minor, also known as the anterior forceps, runs bilaterally and had the lowest reliability estimates for FA (ICC = 0.76, 95% CI [0.61, 0.86]). The test–retest ICCs for all regions of the atlas fell into the "excellent" range; however, a number of the regions had the lower confidence interval overlap with the "good" threshold (Appendix Supplement Table A.6).

Radial diffusivity (RD)

Radial diffusivity has been previously used as a proxy measure for myelin damage or demyelination. In our study, we found the measure to be, on average, of "good" to "excellent" reliability (M = 0.84, 95% CI [0.73, 0.90]). The forceps major, also known as the posterior forceps, was the white matter fiber bundle with the highest RD reliability (ICC = 0.94, 95% CI [0.90, 0.97]). The right hippocampal cingulum bundle had the lowest scan–rescan reliability (ICC = 0.69, 95% CI [0.52, 0.81]). Of note, the lower confidence interval around the ICC was "fair" for five white matter tracts (Appendix Supplement Table A.7).

Mean diffusivity (MD)

Mean diffusivity summarizes the average diffusion properties of a voxel and can be sensitive to pathology such as edema and necrosis, among others. Overall, mean atlas-averaged ICCs were "good" to "excellent" (M = 0.81, 95% CI [0.69, 0.89]). Based on the lower CI bounds, six white matter tracts overlap with the "fair" reliability classification (Appendix Supplement Table A.8). Among those are the bilateral cingulum bundles surrounding the hippocampus (left; ICC = 0.65, 95% CI [0.46, 0.79], right; ICC = 0.63, 95% CI [0.43, 0.77]) and the corticospinal tract (left; ICC = 0.69, 95% ICC [0.52, 0.81], right; ICC = 0.66, 95% CI [0.48, 0.79]). Once again, the highest

scan–rescan reliability was observed in the measurement of the forceps minor (ICC = 0.91, 95% CI [0.85, 0.95]) and forceps major (ICC = 0.93, 95% CI [0.89, 0.96]).

Axial diffusivity (AD)

Axial diffusivity measures water diffusion along the principal axis of diffusion and may be correlated with axonal injury. AD had the lowest average ICC of the DTI scalars in our study (M = 0.78, 95% CI [0.65, 0.87]). While the overall ICC can be classified as "good" to "excellent," there is some regional variability of note (Appendix Supplement Table A.9). The right hippocampal cingulum bundle had the lowest scan–rescan ICC, with the lower confidence interval crossing into the "poor" classification (ICC = 0.53, 95% CI [0.30, 0.70]). Nevertheless, some regions stood out for their excellent scan–rescan reliability, such as the forceps major (ICC = 0.93, 95% CI [0.88, 0.96]) and the bilateral anterior thalamic radiation (left; ICC = 0.90, 95% CI [0.84, 0.94], right; ICC = 0.90, 95% CI [0.83, 0.94]).

4.5 Discussion

In this paper, we report the reliability of nine MRI-derived measures of cortical and white matter morphology and integrity based on 100 scans from 50 youth. Despite the high prevalence of anxiety and ADHD disorders in our young sample, we found good to excellent reliability for all measures. White matter volume was most consistently reconstructed with a scan–rescan ICC of 0.98 averaged across the white matter atlas. Axial diffusivity was the least reliable, with an average ICC of 0.78 across scan sessions. We also observed regional variability in reconstruction, with many structures showing excellent stability across measures, and some showing poor to fair reconstruction. This analysis might be of particular interest for hypothesis driven studies focusing on select regions of interest, and for exploratory and predictive multivariate studies to cross reference the pattern of findings to their reported reliability distributions.

The excellent reliability of gray matter measures should be interpreted in the context of prior work. While the reliability of functional MRI data in youth has received some attention (Thomason et al., 2011; Vetter et al., 2017), literature examining the

reliability of structural MRI data remains sparse. Therefore, we interpret the consistency of our data by comparing it to similar work in adult samples. Iscan and colleagues (Iscan et al., 2015) reported a comparable analysis to ours. Their study included 40 healthy controls (age 18–65), scanned twice, whose MRI images were processed in FreeSurfer. Overall, 25 individuals passed their thorough quality control. In the approved scans, reported ICCs for cortical thickness/ surface area/ volume were 0.81, 0.87, and 0.88; remarkably similar to our values of 0.82, 0.89, and 0.90. The closeness of these values carries two messages: (i) It is possible to collect highly reliable MRI data from young people with anxiety and/or ADHD, and (ii) after proper quality control, the reliability can compare to that attained from scans of healthy adults.

4.5.1 Regional reliability of grey matter

We extended our gray matter analysis to investigate the reliability of cortical folding (LGI), an important neurodevelopmental marker that is essential to the optimization of axonal wiring and the functional organization of the brain (Klyachko & Stevens, 2003). With an ICC of 0.85, cortical folding was of excellent reliability, ranking between measures of cortical thickness and cortical surface area. Cortical folding reliability was slightly lower than what was reported (ICC = 0.94) in a recent paper (Madan & Kensinger, 2017). The difference can be attributed to several factors, as the prior work focused on healthy adults who were scanned either 10 times or with a sequence specifically optimized for brain morphology research. To our knowledge, the current study is the first to report on the reliability of this morphological measure in a pediatric risk sample.

Out of all the cortical gray matter measures, cortical thickness had the lowest ICC overall and had the most structures categorized to be of poor reliability based on their lower bound confidence interval. The average thickness of the cortical mantle is 2.5 mm (Fischl & Dale, 2000) which is close to the 1 mm spatial resolution of most scan sequences. Thereby cortical thickness measurements may be particularly sensitive to motion artifacts even in high-quality data (Alexander-Bloch et al., 2016).

The structures with the least reliable cortical thickness reconstructions were the temporal pole, frontal pole, medial orbitofrontal gyrus, and the entorhinal cortex.

The temporal and frontal poles also exhibited reduced reconstruction consistency in analysis of gray matter volume and surface area, and are known to be problematic in the literature (Klapwijk et al., 2019). The medial orbitofrontal gyrus and entorhinal cortex are localized to the inferior aspect of the brain, and their location makes them particularly affected by susceptibility gradients from air-filled cavities, the bone–tissue interface, and orbital artifacts. However, the orbitofrontal and entorhinal cortices are both essential to fundamental aspects of memory and cognition and have been implicated in a wide range of disorders (Baiano et al., 2008; Rolls & Grabenhorst, 2008). Our results suggest the need for stringent quality control and adequately powered samples in future studies of the cortical thickness of these areas.

4.5.2 Regional reliability of white matter

In contrast, white matter volume had the highest reconstruction reliability in our study. The near-perfect ICC, both regionally and overall, makes the measure particularly suitable for longitudinal research. However, the assessment of white matter microstructure with diffusion tensor imaging (DTI) was more variable. DTI is widely used to infer white matter microstructure, structural connectivity, and axonal health. Our results ranged from good to excellent (ICC 0.78–0.88) for the four DTI measures, with axial diffusivity (AD) being the least reliable and fractional anisotropy (FA) the most reliable. This mirrors the relative interest attained for these measures in the research community. AD may be a correlate of axonal injury (Budde, Xie, Cross, & Song, 2009); however, the measure is less widely used than FA which has been the most popular correlate of white matter integrity (Soares, Marques, Alves, & Sousa, 2013). Regionally, none of the lower confidence intervals for FA ICCs crossed below the good into the fair or poor classification.

Across all DTI measures, the only region with the lower bound confidence interval in the poor classification was the hippocampal cingulum bundle. This white matter tract, along with the cingulum cingulate bundle, had the lowest scan–rescan reliability estimates for AD, MD, and RD. The cingulum bundle is a large white matter tract interconnecting the frontal, parietal, medial temporal, and other areas and has been implicated in a spectrum of neuropsychiatric disorders (Bubb, Metzler-Baddeley, &

Aggleton, 2018). Its size and midline positioning might make it particularly susceptible to motion artifacts and spatial misregistration errors, and thus, a similar warning akin to low reliability areas of cortical thickness applies here as well.

Lower scan–rescan reliability also applied to the corticospinal tract. Interpreting these findings in the context of prior research might be illuminating. Investigations of the underlying reliability of white matter measures in pediatric samples have mainly been restricted to small samples, specific illness, or a limited number of white matter tracts (Alhamud, Taylor, Laughton, van der Kouwe, & Meintjes, 2015; Bonekamp et al., 2007; Carlson et al., 2014). However, a recent paper has examined FA reliability in a well-powered sample comprising of both an adult and an adolescent group (Acheson et al., 2017). Similar to our results, the authors found that in adolescents, the lowest reliability was observed in the corticospinal tract. This observation held in adults, signifying low reliability of the corticospinal tract across development. The corticospinal tract is a white matter motor pathway, and thus, the reliability concerns might not be immediately relevant to psychiatric research.

4.5.3 Comparisons with functional MRI

Lastly, our structural MRI reliability estimates were higher than those reported in functional MRI literature. An early account provided the first empirical evidence of the longitudinal reliability of resting state fMRI in children (Thomason et al., 2011). The authors obtained positive ICC values for the majority of brain voxels, indicating stability within participants across measurements. The first group to investigate the reliability of resting state fMRI in clinical developing groups observed fair (>0.40) to good (>0.70) ICC in the short term (Somandepalli et al., 2015). The authors noted higher ICC in typically developing children compared to those with ADHD. A more recent report examined reliability in adolescent fMRI within a 2-year period (Vetter et al., 2017). The investigators found both variability and stability, with the reliability results dependent on task domain and region of interest. For example, whole-brain ICC was lower (0.44) in cognitive control paradigms and higher (0.74) in reward paradigms. There was great variability across regions of interest, with ICCs ranging from poor (0.19) to excellent (0.84). Two recent meta-analyses suggest that even these modest fMRI reliability values are potentially optimistic.

One meta-analysis examined a decade of test–retest reliability work surrounding functional connectivity. The authors concluded that most functional connections exhibited "poor" ICC of 0.29 (95% CI [0.23, 0.36]; Noble, Scheinost, & Constable, 2019). Another recent meta-analysis examined test–retest reliability of common task-based fMRI measures (Elliott et al., 2020). Echoing the previously mentioned findings, their work revealed poor to fair overall reliability (ICC = 0.40) across 90 studies. However, it is worth noting that contrasting the reliability of structural and functional MRI is not an apples-to-apples comparison. The excellent structural reliability we report in this manuscript is based on the consistent reconstruction of a priori anatomically defined regions. Functional reliability deals with spatial, temporal, and frequency domains that often try to map onto fluid brain processes. Nevertheless, the discrepancy between the two modalities is worth acknowledging as it can have practical applications, such as sample size requirements for biomarker discovery (Elliott et al., 2020).

4.5.4 Limitations

There are several limitations to this study. Sample size is of concern, not in respect to accurately estimating reliability but to problems of scale. Our approach of manual ratings for raw data followed by automated quality assessment for processed data can become resource intensive for large-scale projects, such as the modern biobanks collecting tens of thousands of scans. Our relatively small number of excluded scans would grow substantially in those samples and could potentially vary between groups of interest, for example, those with or without psychopathology. Nevertheless, this is actively being addressed with behavioral interventions before or during scanning, with optimized sequences utilizing prospective motion correction, as well as at the study design phase with oversampling of at-risk youth.

We were also restricted to a single scan site, and data were acquired on the same scanner at all time points. In our study, we found that results generalized to the same scanner over a year later (Appendix Supplement Table A.10). However, large collaborative efforts are often made possible by acquisitions on scanners from different manufacturers at sites that may be continents apart (P. M. Thompson et al., 2014). This can increase variability that confounds the effects of interest. Nevertheless, these

challenges are being overcome with the standardization of scanning parameters and statistical techniques that correct for site differences (J. Chen et al., 2014). Lastly, beyond site differences, variations in data analysis methods are more likely to have a stronger effect on neuroimaging results, but are also being addressed (Nichols et al., 2017).

Another limitation is our choice of parcellation scheme for assessing regional reliability of cortical areas and white matter tracts. The construction of an accurate map of the major subdivisions of the human brain is a century-old endeavor with an accompanying and equally long debate on what constitutes a boundary. There are other parcellations than the one used in this paper that are more biologically grounded, accounting for cortical architecture, topography, and functional connectivity (Glasser, Smith, et al., 2016). However, given that it is impossible to exhaustively test each parcellation, we decided to focus on those most likely to be commonly used in the field. The Desikan atlas has over 5,000 citations on PubMed, and the JHU atlas has almost 2,000. They come default or preinstalled with commonly used MRI software, including FreeSurfer and FSL, respectively. Thus, these atlases are the starting point for a great number of neuroimaging researchers and a basis of comparison for those on the cutting edge who choose to use newer or custom parcellations.

4.6 Conclusions

In conclusion, while researchers should be cognizant of regional variability in reconstruction, pediatric MRI brain data can be highly reliable overall. Furthermore, the high reliability was established in youth at risk for mental illness or those already affected by anxiety and neurodevelopmental disorders. This bodes well for work investigating the neurodevelopmental markers of mental illness at an early stage, before medication, drug use, and other confounds take a persistent toll on the brain. Confidence in the data quality of high-risk youth samples is also a prerequisite for improved diagnosis and development of personalized prevention strategies based on brain markers.

Chapter 5

Deviations from typical brain development are associated with adversity, depression, and poor functional outcomes in at-risk youth

5.1 Abstract

Introduction The majority of psychiatric disorders onset in the first two decades of life. Because a multitude of subtle developmental brain deviations may be associated with risk of mental illness, there is a need for a general index of deviation from typical development. In the present youth study, we examined the extent to which environmental adversity, developmental antecedents, major depressive disorder (MDD), and functional impairment correlate with deviation from normative brain development.

Methods We applied gradient boosting to train a brain age prediction model in a training set of 1299 typically developing youth (age range 9-19 years old, $M = 13.5$, $SD = 3.04$), validated the model in a holdout set of 322 youth ($M = 13.5$, $SD = 3.07$), and used it to predict age in an independent risk enriched test sample ($M = 13.6$, $SD = 2.82$). We tested associations between the brain age gap (predicted - chronological age) and a range of risk factors, psychopathology and functional outcomes associated with mental illness.

Results The mean absolute error (MAE) was 1.53 years in the training set, and generalized to the validation set (1.55 years) and independent at-risk sample (1.49 years). The brain age estimate was highly reliable in repeated scans (intra class correlation = 0.94). Experience of multiple environmental advertises ($\beta = 0.18$, $p = 0.02$, 95% CI [0.04, 0.31]), presence of MDD ($\beta = 0.59$, $p = 0.01$, 95% CI [0.18, 0.99]) and functional impairment ($\beta = 0.16$, $p = 0.01$, 95% CI [0.05, 0.27]) were associated with a positive brain age gap.

Conclusions Several indicators of mental illness had a tendency of reflecting in an older appearing brain. Overall, deviation from normative brain age might be a general indicator of ill brain health.

5.2 Introduction

Most psychiatric disorders onset in the first two decades of life (Kessler et al., 2005; Caspi et al., 2020). The highest risk of mental illness onset coincides with a period of rapid brain development. Cortical grey matter volume shows a non-linear decline during adolescence, explained by reductions in cortical thickness and surface area (Tamnes et al., 2017; Walhovd et al., 2017). Cortical white matter volume increases from childhood until mid to late adolescence (Mills et al., 2016). Aberrations or exaggerations of these typical developmental changes are likely related to the etiology of mental illness (Huttenlocher, 1994; Paus et al., 2008; Whitaker et al., 2016). Thus it is paramount to determine how biological and environmental factors impact developmental trajectories (Casey et al., 2018). Deviation from expected trajectories can be captured with an innovative approach capable of assessing neuroanatomical maturity (Franke, Luders, May, Wilke, & Gaser, 2012; Brown et al., 2012; Chung et al., 2018).

Advances in machine-learning (ML) algorithms combined with access to large databases of brain scans have made it possible to estimate brain-predicted age, or brain age, from MRI images (Cole & Franke, 2017). By analysing scans from typically developing individuals, ML determines the relationship between neuroanatomy and chronological age, across the age range, thereby mapping normative developmental trajectories. Once trained, the ML algorithm is able estimate brain age from previously unseen MRI scans from different individuals and populations. The difference between a person's predicted age and their chronological age results in the brain age gap (Franke & Gaser, 2019). A wide gap suggest deviation from normative brain developmental trajectories and has been associated with a spectrum of ill health.

In adults, increased brain age has been associated with cognitive impairment, traumatic brain injury, Alzheimer's disease, and increased risk of mortality (Cole, Ritchie, et al., 2017; Cole & Franke, 2017; Franke & Gaser, 2019). A higher brain age, suggestive of an older-appearing brain, has also been associated with severe mental illness (SMI), including both mood (Han et al., 2020) and psychotic disorders (Schnack et al., 2016). A recent analysis of over 40,000 scans has revealed that all clinical groups, including those with major depressive disorder (MDD), bipolar disorder (BD) and schizophrenia exhibited a positive brain age gap when compared to healthy controls (Kaufmann et al., 2019). Thus, while the brain age literature is

well established in studies of adults, less is known about how the brain age gap relates to mental illness in adolescence.

Prior work has established the validity of brain age prediction in the developmental context, with the ability to reliably differentiate between age groups (childhood, early adolescence, middle adolescence, late adolescence) (Franke et al., 2012; Brown et al., 2012). The brain age gap has been related to cognitive performance in typically developing youth (Lewis, Evans, & Tohka, 2018). However, applications in mental health contexts have been more limited. The few available adolescent brain age studies have shown an increased brain age gap associated with MDD (de Nooij et al., 2019) or risk for conversion to psychosis (Chung et al., 2018). Building on the prior literature, we wanted to examine if the brain age gap may work as a general biomarker of deviation from typical development associated with multiple adverse exposures and outcomes.

In the present study we built a model optimized for predicting age in typically developing youth. We then applied the model in a well characterized high-risk sample of youth in order to calculate the brain age gap for each individual. Considering that there is no single indicator of ill mental health in the developmental context, we implemented a multi-prong approach (Polanczyk et al., 2010). First, we examined the effect of multiple negative environmental factors, such as low socio-economic status and maltreatment on predicted brain age. Second, we investigated the effect of early transdiagnostic symptoms for mental illness. Third, validating prior work, we examined if presence of MDD was associated with increased brain age. Lastly, we examined the effects of functional impairment on brain age. Given prior evidence, we expected that all indicators would non-specifically reflect in a positive brain age gap, signifying a global factor of ill-being.

5.3 Methods

In this study we implemented a testing, validation, and application strategy. In order to model normative development, we acquired scans from typically developing youth from neuroimaging databases. The scans were then partitioned into a training set, for building the model, and validation set, for assessing generalized performance. An independent cohort, enriched for risk for mental illness, was used for hypotheses testing. The inclusion age criterion was 9 to 19 in order to capture development during adolescence.

5.3.1 Sample

We leveraged publicly available MRI data in order to build multivariate models for predicting chronological age of typically developing youth. A summary of the datasets, all of which have been used in prior publications, is available in Table 5.1. Each cohort collected data with the participants' informed consent and approval by local institutional review and ethics boards. Inclusion criteria for the typically developing cohorts was participants ages 9 to 19 years old, with IQ above 75 where applicable, passing automated quality control. Figure 5.1 shows sample size, age, and sex distribution of included cohorts.

We then applied the best model to an independent sample of youth at risk for mental illness (FORBOW), Demographics Table 5.2. FORBOW exclusion criteria were personal history of 1) psychotic illness or autism spectrum disorder, 2) any serious medical or neurological disorders, or 3) MRI contraindications and 4) Full scale intelligence quotient (FSIQ) below 75 on the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011).

Data partition

Training set 80% (1299) of the typically developing controls were used for model training.

Validation set 20% (322) held out controls were used for model validation and estimation of model bias (Smith, Vidaurre, Alfaro-Almagro, Nichols, & Miller, 2019).

Table 5.1: Description of the cohorts

Cohort	Description	Website	Age ± SD
Autism Brain Imaging Data Exchange (ABIDE)	ABIDE is a product of an international multisite collaboration, openly sharing neuroimaging data. We only included typically developing controls (Craddock 2013)	http://preprocessed-connectomes-project.org/abide/	13.38 ± 2.49
ABIDE II	ABIDE II has aggregated a significant number of additional datasets to promote discovery science (Di Martino 2017)	http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html	11.64 ± 2.34
Child Mind Institute: Healthy Brain Network (CMI)	The CMI has launched an ongoing initiative focused on collecting and sharing a large biobank of multimodal brain imaging data from New York City area youth (Alexander 2017)	http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/sharing_neuro.html	12.64 ± 2.67
Consortium for Reliability and Reproducibility (CoRR)	CoRR was created as an open science resource focused on test-retest reliability of MRI connectivity work (Zuo 2014). The data, including structural images, were gathered across the world and shared through the International Neuroimaging Data-sharing Initiative (INDI)	http://fcon_1000.projects.nitrc.org/indi/CoRR/html	15.39 ± 3.13
The NIH MRI study of normal brain development (NIH)	The NIH pediatric study utilized a uniform acquisition protocol and epidemiologic sampling to form an MRI database of typically developing children (Evans 2006). The data was shared with researchers and the clinical medicine community in the mid 2000s	https://www.bic.mni.mcgill.ca/nihpd_info/info2/data_access.html	13.26 ± 2.80
Pediatric Imaging, Neurocognition, and Genetics (PING)	PING was tasked in creating a large repository of behavioural, imaging, and genetics measures from typically developing children (Jernigan 2016)	https://chd.ucsd.edu/research/ping-study.html	13.49 ± 2.92
Families Overcoming Risks and Building Opportunities for Well-being (FORBOW)	FORBOW is a longitudinal study enriched for sons and daughters of parents with mental illness (Uher 2014). Behavioural and psychopathology data collected 2013-2020 and brain measurements completed in 2016-2020	http://www.forbow.org	13.57 ± 2.82

Table 5.2: Demographics of risk enriched FORBOW cohort.

SMI = Severe Mental Illness, MDD = Major Depressive Disorder, ADHD = Attention-Deficit/Hyperactivity Disorder FSIQ = Full Scale Intelligence Quotient. SES = Socioeconomic status (larger number represents higher SES)

	n (%)	***n scans***
Female	79 (53%)	187
Family history of SMI	118 (68%)	248
Lifetime MDD	24 (16%)	64
Lifetime Anxiety Disorder	63 (42%)	159
Lifetime ADHD	37 (25%)	96
Any cannabis use	15 (10%)	43
	Mean SD	
FSIQ	106 13.5	
SES (0-5)	3 1.42	

Test set We tested hypotheses in a separate at-risk cohort (FORBOW), fully independent from training and validation. This included 338 scans from 150 individuals (see Supplemental Table A.11 for scan timing breakdown).

5.3.2 FORBOW assessments

Polyenviromic risk score (PolyE)

The adversity score was calculated as a mean of 10 binary indicators of socio-economic adversity and victimization: (1) biological mother's education, (2) biological father's education, (3) home-ownership status, (4) annual household income, (5) emotional abuse, (6) physical abuse, (7) sexual abuse, (8) neglect, (9) exposure to violence at home and (10) bullying. Our prior work goes into greater detail on these indicators (Zwicker, MacKenzie, et al., 2019). Considering that there is no psychometrically derived cut-off for this measure, in instances where binarizing for visualization was necessary PolyE was considered high if there was any childhood maltreatment.

Youth Experience Tracker Instrument (YETI)

The YETI is a brief self report measure of six antecedents that precede and predict mental illness: affective lability, anxiety, depressive symptoms, basic symptoms, psychotic-like experiences, and sleep (Patterson et al., 2020). The questionnaire was

designed to facilitate early identification of risk for severe mental illness and has been validated in the FORBOW cohort. A score of ≥ 8 indicates high antecedent burden.

Major depressive disorder (MDD)

Participating youth were interviewed using the Kiddie Schedule for Affective Disorders and Schizophrenia, Present and Lifetime Version (K-SADS-PL; Kaufman et al., 1997) by assessors blind to parent psychopathology. Diagnoses were confirmed in consensus meetings with a psychiatrist.

Columbia Impairment Scale (CIS)

The CIS is a 13 item scale that provides a global measure of impairment (Bird, Shaffer, Fisher, Gould, & et al, 1993). The questionnaire taps into a number of functional domains, including relationships with family at home, relations with peers, academic functioning, and involvement in general hobbies/interests. A score ≥ 15 indicates clinically significant functional impairment (Attell, Cappelli, Manteuffel, & Li, 2018). We averaged both child and parent reporting on the CIS.

5.3.3 MRI acquisition

Images were acquired with a 3T General Electric Discovery MR750 scanner equipped with a 32-channel MR Instruments RF head coil. Scanning took place at the Biomedical Translational Imaging Centre (BIOTIC), Halifax, Nova Scotia. We collected a 3D T_1-weighted (T1w) Brain Volume imaging (BRAVO) sequence with whole-brain coverage, 1 mm^3 isotropic resolution, matrix = 224 $\times$ 224, field of view (FOV) = 224 mm, 168 sagittal slices at 1 mm thickness, repetition time (TR) = 5.9 ms, echo time (TE) = 2.2 ms, inversion time (TI) = 450 ms, flip angle = 12°. In addition, we collected a T_2-weighted fluid attenuated inversion recovery (FLAIR) sequence using a T2 prep contrast option (T2PREP) with identical coverage, resolution and acquisition orientation to the T1w sequence, TE = 98 ms, TR = 5,100 ms, TI = 1,427 ms, echo train length (ETL) = 250 echoes, flip angle = 90°.

5.3.4 MRI processing and features

Data across all cohorts was processed with FreeSurfer 6 software (RRID:SCR_001847; Fischl, 2012); *FreeSurfer website*. In depth reporting of the processing in the FOR-BOW sample as well as assessment of MRI reliability is available in our prior work (Drobinin, Gestel, et al., 2020); *GitHub repository*.

Automated quality assurance was done with the Qoala-T tool (Klapwijk et al., 2019); *Qoala-T website*. Qoala-T is machine learning tool designed to automatically classify the quality of FreeSurfer output. We report on data that was recommended for inclusion by Qoala-T.

For cortical features to be used in the machine learning analysis, we compiled 34 cortical structures per hemisphere based on the FreeSurfer default Desikan–Killiany atlas (Desikan et al., 2006). We included cortical gray matter volume and cortical surface area measurements for each structure. We excluded cortical thickness measurements for several reasons: (1) Cortical thickness measurements are the least reliable and most sensitive to MRI artefacts (Drobinin, Gestel, et al., 2020; Iscan et al., 2015), (2) cortical thickness measurements are more scanner specific than other measures (Fortin et al., 2018) and there is no agreed upon way to standardize them within a cross-validation framework, (3) we wanted to maintain an adequate ratio of features/predictors to sample size. Overall, we included 136 cortical features.

We also included bilateral global (e.g. intracranial volume) and subcortical measures (e.g. amygdala volume) from the FreeSurfer output. This totalled 53 additional structures, for an overall total of 189 features for use in training (see Supplement Table A.12 for full feature list).

5.3.5 Statistical analysis

All statistical analyses were performed in R Studio (R Version 3.6.3; RStudio version 1.3.959; RStudio Team, 2019). Machine learning was performed within the tidymodels framework (Kuhn & Wickham, 2020). Tidymodels (version 0.1.1) is a standardized collection of packages for modelling and machine learning in R: tidymodels website. Associations with clinical variables were analysed with mixed effect linear models implemented in the lme4 package version 1.1.23 (Bates, Mächler, Bolker, & Walker, 2015).

Data pre-processing and cross-validation The volume of the 5th ventricle was removed due to near zero variance, reducing the feature set to 188. We implemented 10 fold cross-validation, repeated 10 times, with the folds stratified by the scan age.

Model specification and tuning The model setup was to predict scan age using cortical and subcortical brain features. We implemented the XGBoost machine learning algorithm version 1.0.0.2 (T. Chen & Guestrin, 2016) due to its top performance in many machine learning challenges as well as successful use in the largest brain age study to date (Kaufmann et al., 2019). The primary performance metric was to minimize mean absolute error (MAE), expressed in years.

The model was specified with the parsnip package from the tidymodels framework. The XGBoost model was set to regression mode, with 1500 trees. The following hyperparameters were tuned through cross validation: tree_depth, min_n, loss_reduction, sample_size, mtry, and learn_rate. Package specific nomenclature for these parameters is available in Supplemental Table A.13. Tuning was done with grid search (latin hypercube sampling, size = 500). The method involves near-random sampling of parameter values from a distribution of all possible parameter values.

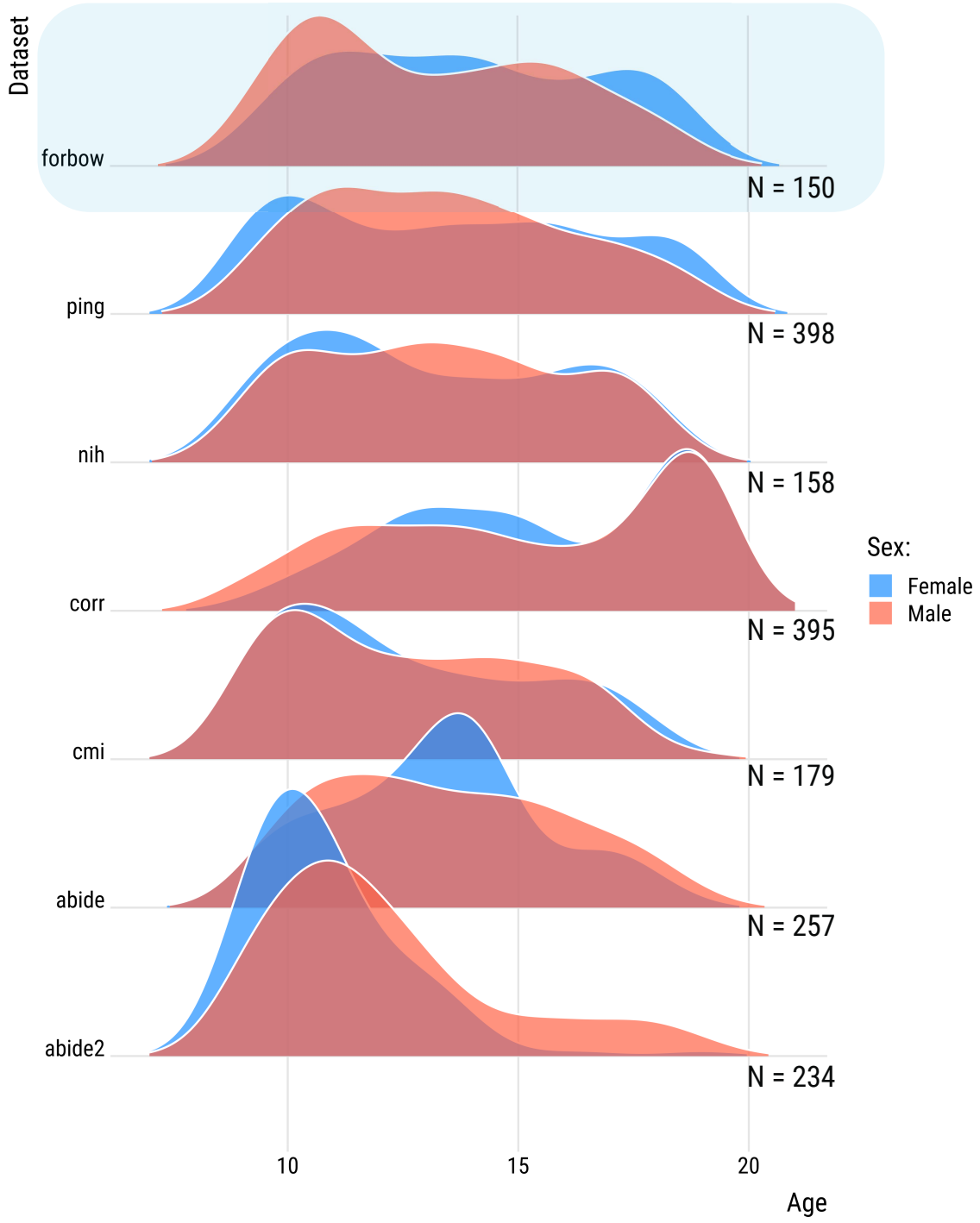
Finalizing and visualizing model From the tuned cross validated results, we selected the best model (lowest MAE) using the "one standard-error rule" (Breiman, Friedman, Stone, & Olshen, 1984). In other words, we picked the most simple model, in terms of tree depth, that is within one standard error of the best performing model. The depth of the tree represents the degree of feature interaction, and can lead to high non-linearity and variance (James, Witten, Hastie, & Tibshirani, 2013). The best model was then finalized on the whole training set and predictions were made on the validation set and on the FORBOW test set. The final fitted model is freely available for researchers to make predictions on their own developmental data: *GitHub link*.

Neuroanatomical contribution to age prediction was visualized using the variable importance plots (vip) R package version 0.2.2. The resulting importance of specific brain regions in predicting age was plotted with the ggseg R package version 1.5.4 (Mowinckel & Vidal-Piñeiro, 2019).

Brain age prediction and bias correction The finalized XGBoost model was used to predict scan age in the validation and test sets. We also computed the brain age gap by subtracting the true scan age from the predicted age. Across regression methods there is a tendency for prediction bias towards the group mean (Smith et al., 2019). In the context of brain age studies this is seen as a slight age overestimation in younger participants and underestimation of predicted age in the older participants. To adjust this, we used a linear bias correction method proposed by prior research (de Lange & Cole, 2020; Smith et al., 2019). We fit a linear regression model on the validation set: $predicted\_age = \alpha + \beta_1(scan\_age) + \epsilon$. To attain a bias corrected prediction we subtracted the intercept from the predicted age and then divided it by the slope. The slope and intercept are generalizable when applied to new data (Peng, Gong, Beckmann, Vedaldi, & Smith, 2019), thus we used these coefficients from the validation set to correct the bias in the test set without requiring any information about the independent test sample.

Hypotheses testing We used mixed effect linear models to determine the association between the corrected brain age gap and the variables of interest. We ran separate models with the PolyE, YETI, MDD, and CIS as the primary independent measures and the brain age gap as the dependent measure. As recommended by prior work (Le et al., 2018), we covaried for the effects of age, age^2 and sex in the main analyses. For sensitivity analysis we additionally covaried family history of SMI, socio-economic status (SES), any prior cannabis use, and IQ. We modelled multiple observations from each participant by including the scan id as a random effect. We also accounted for the non-independence of brain data from related individuals by including the family identifier as a random effect. We reported effect sizes using standardized regression estimates (β) and their 95% confidence intervals (CIs).

Figure 5.1: Age and sex distributions of all datasets. Risk enriched FORBOW sample highlighted above.



5.4 Results

5.4.1 Brain age model performance

Training performance We built a brain age model using magnetic resonance imaging scans of typically developing youth from 6 cohorts (n = 1299; age range 9-19 years old). The top model had an MAE of 1.53 years in cross-validation. See Supplement Table A.14 for top 10 models and their parameters.

Validation performance The model performed with an MAE of 1.55 years on the validation set, similar to the cross-validation performance. As expected, the brain age gap was correlated with age, showing bias ($r(320)$ -0.73, $p < 0.001$). We assessed the relationship between the predicted age and the chronological age: $predicted\_age = 6.41 + 0.55(scan\_age) + \epsilon$. As recommended in the literature, we subtracted the intercept from the predicted age and then divided it by the slope for bias correction. This procedure inflated the variance of the predicted age, increasing the MAE to 1.98, however it eliminated the correlation between the brain age gap and chronological age ($r(320) < 0.001$, $p = 1$). See Figure 5.2 for model performance across datasets.

Test performance Our target of interest was the independent FORBOW cohort, where the unmodified model performed with an MAE of 1.49 years. We used the coefficients from the validation set to bias correct the predicted age. After bias correction the MAE was 1.86 and the brain age gap did not significantly correlate with chronological age ($r(336)$ = -0.07, $p = 0.2$).

5.4.2 Reliability of predicted age

While the reliability of individual brain imaging measurements has been established, the short-term reliability of brain age estimation is unknown. In 50 individuals re-scanned within weeks of their first scan the intraclass correlation coefficient (ICC 1,1) for predicted brain age was 0.94 (95% CI 0.90, 0.96) indicating excellent reliability.

Figure 5.2: Predicted brain age and chronological age at time of scanning in validation (20% holdout set) and independent test set (FORBOW cohort).

MAE = mean absolute error, corrected MAE = bias corrected MAE.

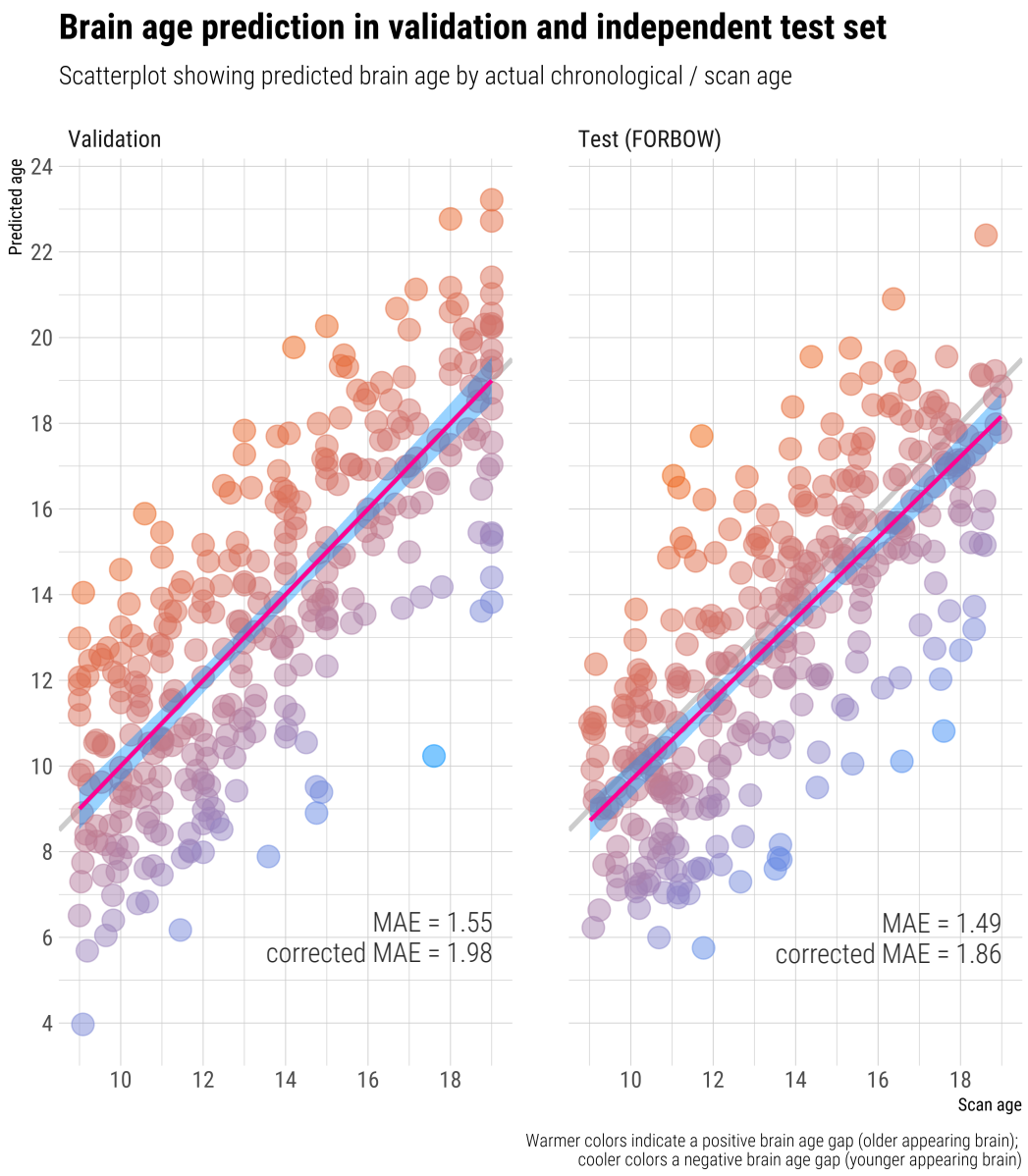
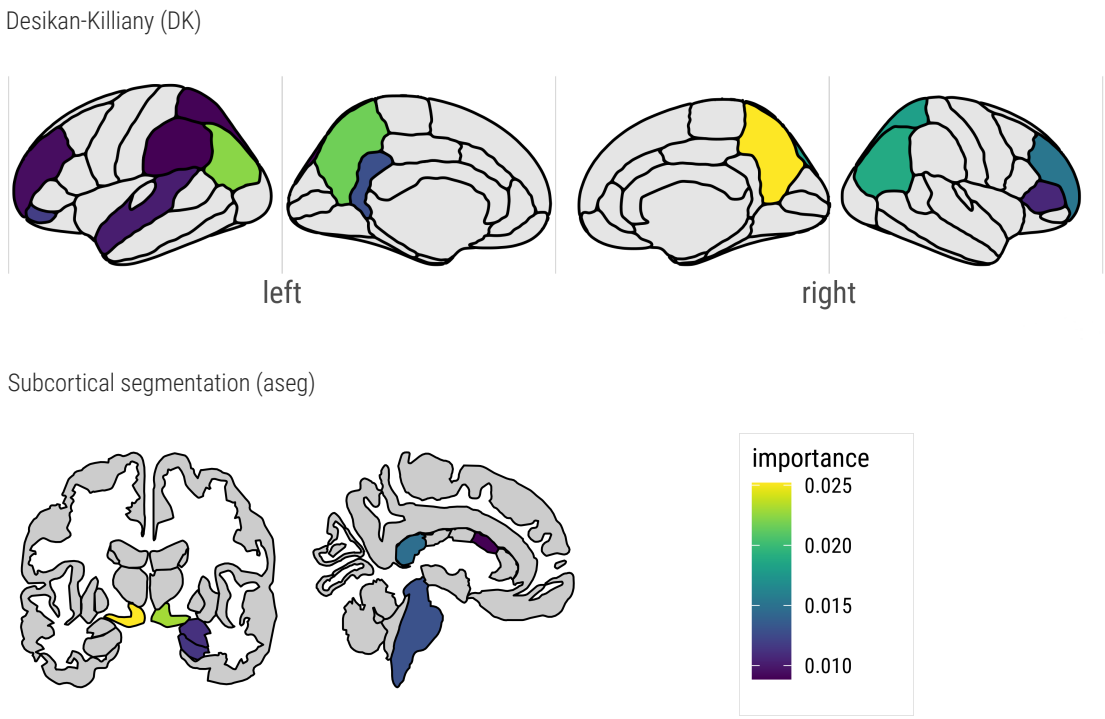


Figure 5.3: Top cortical grey matter volume and subcortical contributions to age prediction.



5.4.3 Neuroanatomical contribution to age prediction

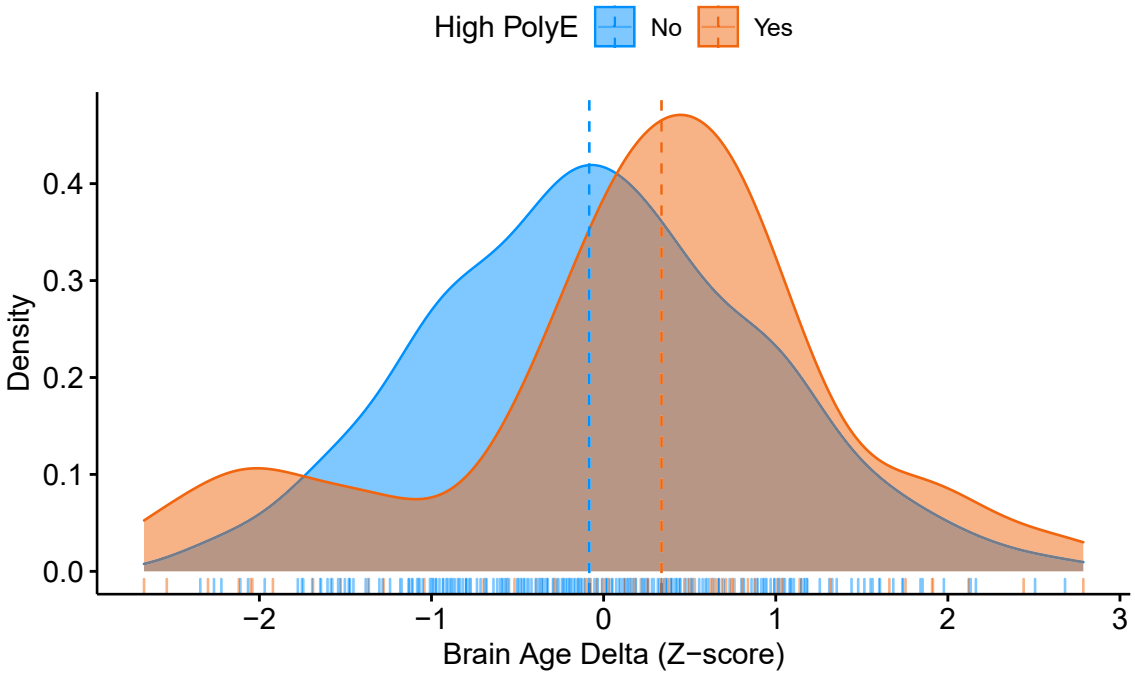
The machine learning model relied on a diverse set of features to make accurate age prediction. Among them were global measures, such as brain segmentation volume, information from the ventricles, parietal regions and a number of other cortical volumetric measures, as well as subcortical data including the hippocampus and the brainstem. Figure 5.3 shows the top grey matter volume and subcortical contributions in the prediction model. Supplemental Figure A.3 lists the relative contribution to brain age prediction of the top 30 structures.

5.4.4 Association with exposures and outcomes

PolyE First we examined how an aggregate measure of multiple adverse environmental events (PolyE) relates to brain development. After controlling for age, age^2 and sex we found that PolyE was significantly associated with a positive brain age

gap ($\beta = 0.18$, $p = 0.02$, 95% CI [0.04, 0.31]). In other words, adversity is associated with older looking brains (Figure 5.4). This effect remained after additionally controlling for the effects of family history of SMI, any lifetime cannabis use, and IQ ($\beta = 0.17$, $p = 0.02$, 95% CI [0.03, 0.31]).

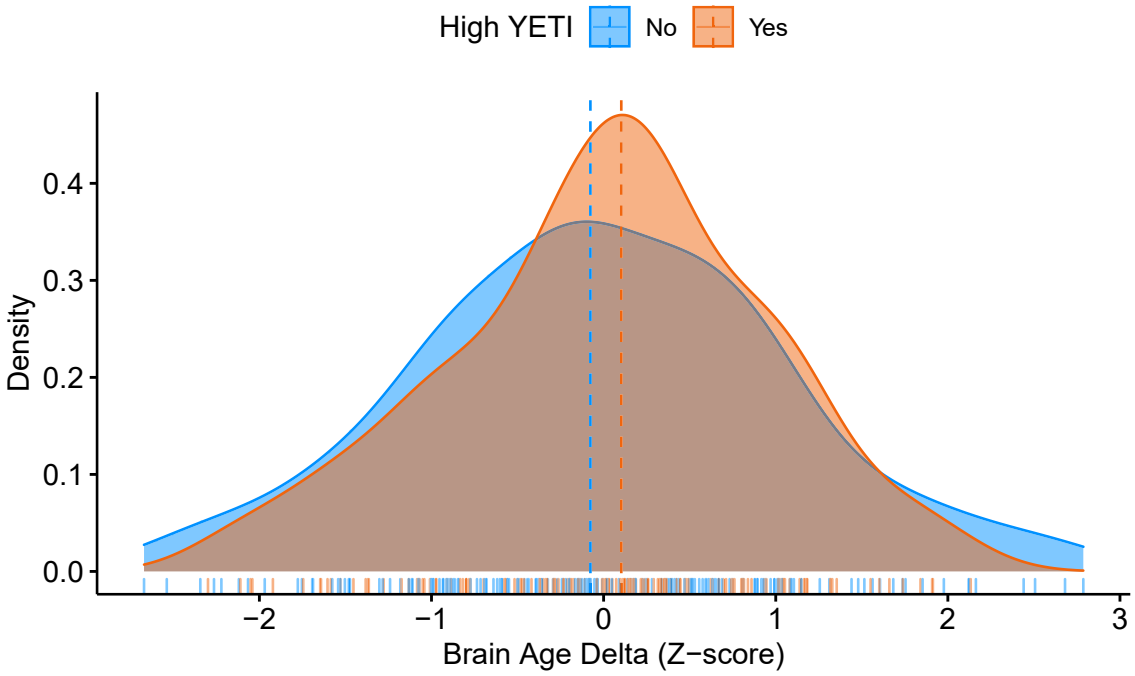
Figure 5.4: Brain age gap as a function of adverse environmental events (PolyE).



YETI Next we tested the association between the Youth Experience Tracker Instrument (YETI) and the brain age gap. We found that the dimensional measure of antecedent burden was not significantly associated with deviation from expected brain age ($\beta = 0.10$, $p = 0.09$, 95% CI [-0.01, 0.21]). Individuals high on the symptom scale were shifted to the right of the brain age gap distribution, towards an older appearing brain, however there was substantial overlap 5.5.

Major depressive disorder We found that MDD was associated with increased positive brain age gap ($\beta = 0.59$, $p = 0.01$, 95% CI [0.18, 0.99]). The effect of an older appearing brain in MDD (Figure 5.6) held when further controlling for family

Figure 5.5: Brain age gap as a function of high number of symptoms on the YETI scale (antecedent burden).

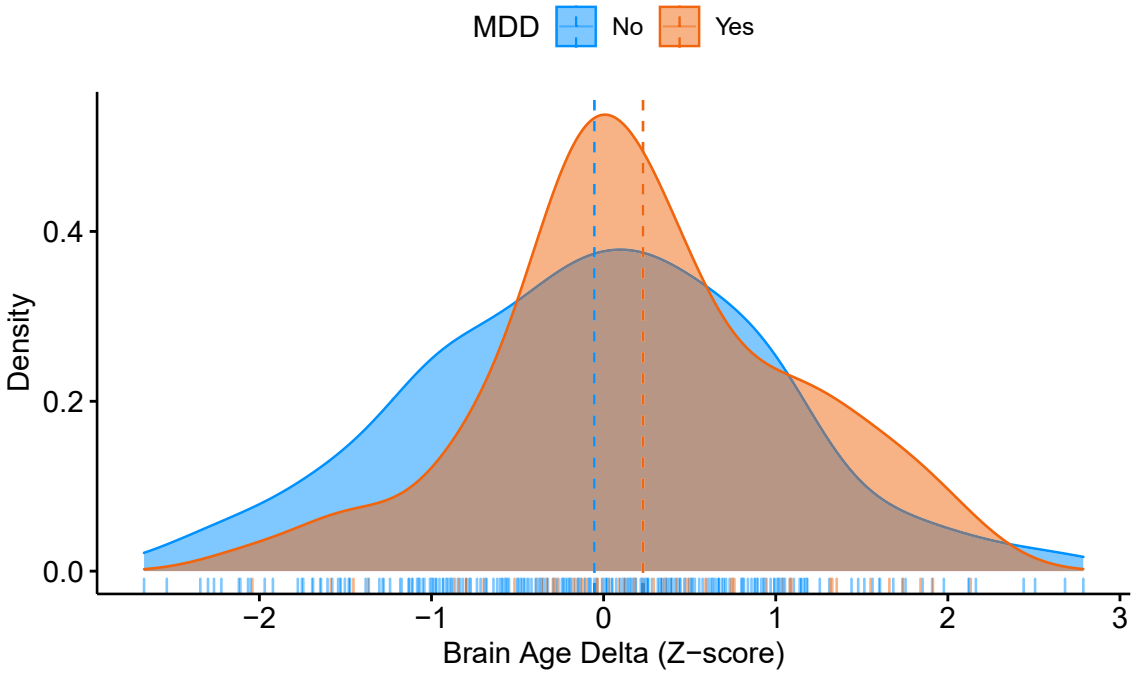


history of SMI, socio-economic status, any lifetime cannabis use, and IQ ($\beta = 0.42$, $p = 0.05$, 95% CI [0.01, 0.84]).

CIS Finally, we examined the effect of functional impairment on the brain age gap (Figure 5.7). The main analysis showed that increasing functional impairment relates to higher predicted brain age ($\beta = 0.16$, $p = 0.01$, 95% CI [0.05, 0.27]). The effect held in the subsequent sensitivity analysis ($\beta = 0.13$, $p = 0.02$, 95% CI [0.02, 0.24]), see Supplemental Figure A.4 for forest plot of effects across models.

High total loading It has been suggested that a small segment of the population carries a large burden in reference to a number of negative outcomes (Caspi et al., 2016). In an exploratory analysis we wanted to examine this effect in our cohort by selecting the individuals that score in the top 20% on the three dimensional measures. We did not find a statistically significant effect of this total high loading ($\beta = 0.37$, $p = 0.12$, 95% CI [-0.09, 0.83]). However, our sample size was underpowered for this

Figure 5.6: Brain age gap as a function of Major Depressive Disorder.



analysis, as thankfully a minority of individuals met these joint criteria ($n = 18$). Nevertheless, Figure 5.8 shows that this trend might warrant future attention.

Figure 5.7: Brain age gap as a function of functional impairment (CIS).

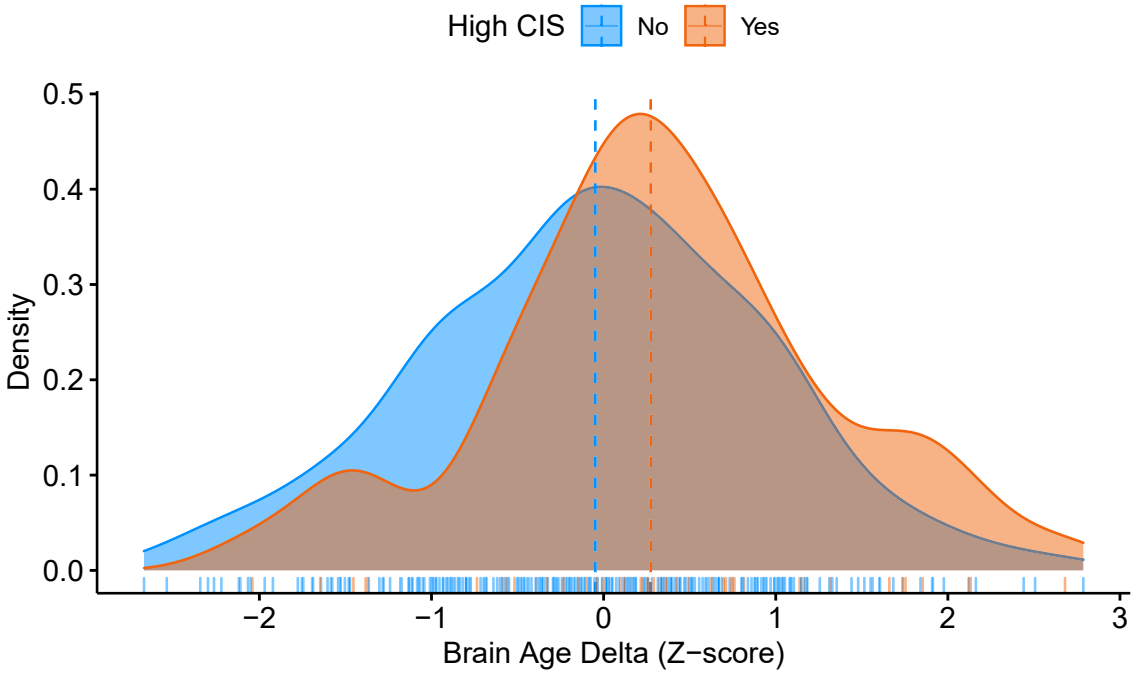
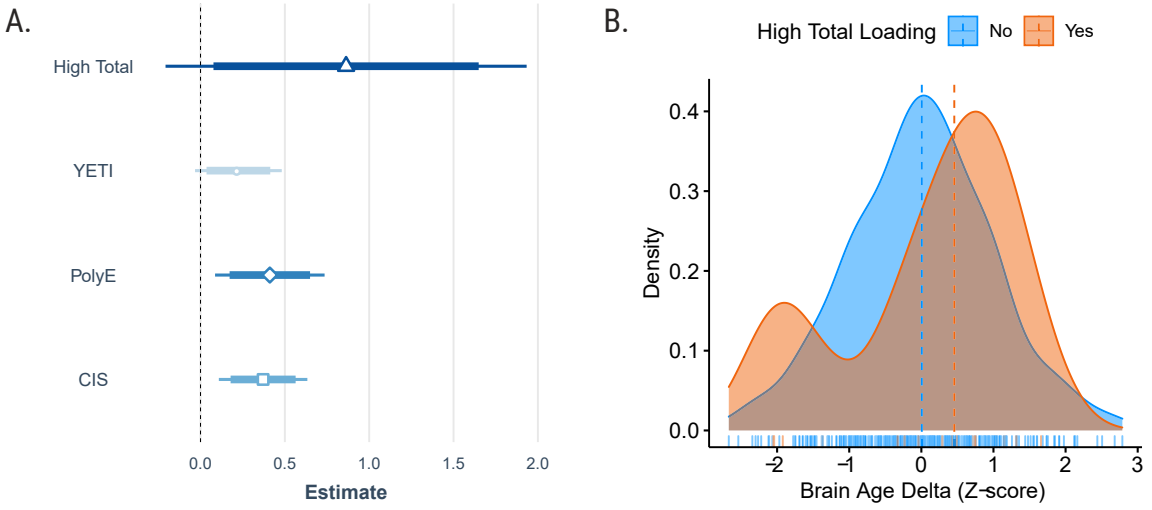


Figure 5.8: Brain age gap as a function of scoring in the top 20% on 3 measures of antecedents, environmental adversity, and functional impairment. A. Forest plot with summary of effect sizes across measures. 95% confidence interval lines are presented with an inner CI of 85%. B. Density plot of brain age gap as function of high total loading.



5.5 Discussion

In this work we were able to accurately and reliably predict age using neuroanatomical data from a large sample of typically developing youth. The prediction accuracy generalized to an independent cohort of youth enriched for risk of mental illness (FORBOW). Within FORBOW we found that a number of factors linked to the causes and consequences of mental illness are associated with discrepancies between predicted and chronological age. The spectrum of adverse experiences, from early antecedents for mental illness, to environmental adversity, adolescent onset depression and functional impairment, all had a tendency of reflecting in an older appearing brain. These effects were present against the backdrop of profound neurodevelopmental change accompanying adolescence. Overall, deviation from chronological age might be a general indicator of unwellness rather than a marker of a specific exposure or outcome.

Our accuracy in predicting chronological age falls in line with prior developmental work with MAEs from one to two years (see Franke & Gaser, 2019 for comprehensive review of the last decade of brain age studies). We were able to achieve a one and a half year MAE while maintaining consistent scale, generalizability and reliability. To our knowledge we have aggregated the largest developmental sample in the study of brain age and leveraged it to show consistent performance between training, validation and independent testing. We did not observe any increase of the prediction error when the model was applied to a cohort of youth enriched for risk of mental illness. Familial risk did not have a significant effect on prediction. Our results show that in the interplay of biology and circumstance, deviation from chronological age more strongly reflects what happens throughout life rather than genetic contributions present from the beginning.

Many individuals experience adversity in childhood and adolescence. In line with our hypothesis, we found that cumulative exposure to adversity was associated with advanced brain age. This novel approach to capturing aggregate adversity can be interpreted in the context of existing neuroimaging literature on the underlying components of the dimensional measure. Community disadvantage and lower socio-economic status have previously been associated with reduced cortical tissue volume, both globally and regionally (Gianaros et al., 2017; McDermott et al., 2019). Our measure also

assessed childhood maltreatment, including physical, sexual, emotional abuse and neglect. Maltreatment has been shown to have lasting effects on the stress response system and neurobiological development (Bremner & Vermetten, 2001; Teicher et al., 2003). The most prominent neuroanatomical effects include grey matter reductions in fronto-limbic structures such as the prefrontal cortex and hippocampus, enlargements in the ventricles, and reductions in the size of the corpus callosum (Hart & Rubia, 2012; Teicher et al., 2004). Considering that typical development from late childhood to early adulthood involves widespread decreases in cortical volume (Mills et al., 2016; Tamnes et al., 2017), the machine learning model is likely sensitive to the additional decreases associated with adversity, resulting in a prediction of advanced maturation, or a higher brain age. Correspondingly, disproportionate increases in measures such as ventricular volumes, which increase with chronological age, would also be associated with higher brain age.

We also hypothesized that developmental antecedents for mental illness would be associated with increased brain age gap. While higher levels of antecedents shifted the distribution towards increased brain age, the effect was not statistically significant. This was unexpected because we previously found that the basic symptoms and psychotic-like experiences captured by the YETI were associated with reduced cortical folding in this cohort (Drobinin, Van Gestel, et al., 2020). It is possible that early pre-prodromal symptoms are associated with subtle cortical alterations that are masked amid the developmental changes occurring in adolescence. Our prior work found reductions in the orbitofrontal and occipital areas, none of which feature prominently in the list of structures that contribute most to accurate age prediction (Supplement Figure A.3). Our current findings compare to those from Chung et al., 2018, where prodrome symptoms were associated with increased brain age, but were not robust enough to predict conversion to illness.

In our final analyses we examined the relationship between depression and the brain age gap, and concluded the investigation by demonstrating the functional relevance of the brain age gap beyond diagnostic group differences. In the present study, MDD was associated with increased brain age, confirming our hypothesis. This finding was supported by a recent, well powered study of increased brain age in adults with MDD (Han et al., 2020). Another recent longitudinal study has shown that for

youth with MDD compared to controls, brain age deviated from chronological age at follow up (de Nooij et al., 2019). Finally, functional impairment is a necessary criterion to meet threshold for many mental health diagnoses and is an important outcome measure in its own right (Attell et al., 2018). Our results showed that functional impairment, as captured by a widely used youth impairment scale, is associated with deviation from chronological age. In one of the largest neuropsychiatric studies to date, Kaufmann et al., 2019, have demonstrated that increased brain age was associated with several disorder specific functional measures. Our study expands on this finding by demonstrating a more generalized effect of functional impairment on brain age during adolescence.

5.5.1 Limitations and future directions

There were several limitations to this study. Prediction error can be further reduced in a number of ways. For example, some scan sites only provided ages as whole integers, imposing a built in cap on prediction accuracy. Furthermore, state of the art deep learning methods, specifically convolutional neural networks (CNN) have shown greater accuracy than offered by more "shallow" machine learning techniques in adult samples (Cole, Poudel, et al., 2017). An additional benefit of CNNs would include less processing requirements for MRI data with traditional software. Further still, we focused on a broad range of cortical and subcortical measures. However, the feature set can be expanded to include multimodal data for improved accuracy and perhaps more importantly mechanistic inference (Engemann et al., 2020).

The brain age gap addresses two major challenges and opportunities in neuroimaging: high dimensionality of brain data and individual level prediction. Regions across the brain are analysed in aggregate through a multivariate framework in order to make accurate individual-level predictions. This process is then summarised by an intuitive single number; the brain age gap. However this reduction in complexity is a double edged sword when it comes to identifying the mechanisms responsible for brain age deviation. Our efforts at interpretable machine learning offered an explanation for which structures contributed most to brain age prediction. By incorporating diffusion, myelination, and data from other modalities future studies may be able

to capture different aspects of pathophysiology contributing to brain age gaps. Perhaps the best example of such work was done by Brown et al., 2012. The group examined age-varying contributions of different imaging measures to the prediction of age in healthy individuals. For example, diffusion measures from white matter fiber tracts were found to be most relevant between the ages of 12 to 15. Improvements on such study designs, done at a population neuroscience level, might offer mechanistic insights into the pathophysiology of brain age deviation in the context of neuropsychiatric disorders.

A new generation of large-scale studies (Alexander et al., 2017; Casey et al., 2018) investigating adolescent brain development across health and disease might provide the data needed to address some of the mentioned limitations. Furthermore such longitudinal biobanks might offer enough power to examine smaller groups that show distinct developmental trajectories. For example, in our study individuals scoring highly on measures of adversity, antecedents and impairment hinted at a particularly strong deviation from chronological age. However the subsample was very small and the exploratory analysis was not statistically significant. Future work can target such subsamples and others in further detail with adequate power to jointly examine the effects of lifestyle factors and medical conditions that are also known to have an effect on brain health (Smith et al., 2019).

5.5.2 Conclusion

In conclusion, we have shown that several factors, including adversity, early antecedents for mental illness, adolescent onset of depression, and functional impairment, were associated with a higher brain age gap in youth. In the spirit of open science we have shared our brain age algorithm with the wider scientific community and offered avenues for future studies of neurodevelopmental trajectories.

Chapter 6

General discussion

In this section I will summarise my main findings and discuss broader aspects of how they fit in the context of current perspectives in neuroimaging.

6.1 Inferior frontal gyrus in BD

In Chapter 2 (Drobinin et al., 2019), I replicated and extended the finding of structural changes in the right IFG volume as a neuroanatomical marker of familial susceptibility for BD. Using 3D representations of the cortical sheet I found that the volumetric enlargement was linked to the increased cortical SA, and not CT or cortical folding. Moreover, I localized the largest volume and SA differences to the pars triangularis of the rIFG, providing a more specific marker of familial risk for BD.

This is a very specific finding: right hemisphere, prefrontal cortex, inferior frontal gyrus, pars triangularis, cortical surface area. However, the way in which it may be referenced has a surprising level of flexibility. Hong, Yoo, Han, Wager, & Woo, 2019, brought light to this issue by highlighting how flexibility in testing spatial hypotheses allows presenting any result as a replicated finding. The authors found that almost half of the 'replicated' findings had spatial co-ordinates more than 15 mm away from the 'original' findings, which is enough to cross boundaries between structures. This is also very relevant for functional MRI research. For example, literature describing 'canonical cognitive control networks' have sometimes displayed vastly different spatial activation maps for those networks. The field would benefit from researchers, myself included, being more mindful of the specificity of prior literature when providing supporting or contrasting evidence.

Nevertheless, this chapter showed the unique nature of rIFG morphology in BD, with larger volume and SA early in the course of illness, providing a new insight into the localization and topology of a prior finding.

6.2 Cortical folding and psychotic symptoms

In Chapter 3 (Drobinin, Van Gestel, et al., 2020), I wanted to determine if youth with psychotic symptoms displayed cortical aberrations before the onset of impairing psychotic illness. Thus I examined 3D reconstructions of cortical folding, an early neurodevelopmental marker of cortical expansion. I found a pattern of decreased cortical folding in adolescents who had psychotic symptoms but who did not meet the criteria for a psychotic disorder.

The results of this study mirror the results of many studies in psychology and neuroscience, in that effect sizes are small and the groups overlap substantially. Small effect sizes face a challenge from the commonly used mass-univariate approach to statistical analysis. In essence, the model of interest is tested for each brain structure in turn, then thresholded by the number of tests done. In the end a few regions might show statistically significant associations with the clinical variable or group of interest. To illustrate an example, Schmaal et al. (2016) examined subcortical brain alterations in MDD. While the study had thousands of participants, only the hippocampal volume differences were significant after correcting for multiple testing.

Returning to my study, after correcting for multiple testing we found only a few differences in frontal and occipital areas. However, if we revisit Figure 3.1, we see a general trend of lower cortical folding in youth with psychotic symptoms across *all* examined regions. Rather than discarding these data, there has been growing interest in better utilizing the joint contributions of small effect sizes (Woo, Chang, Lindquist, & Wager, 2017), such as with mutivariate methods I employed in Chapter 5. Such work has the potential to bring deeper insights into the brain changes shared across disorders.

6.3 Reliability of developmental MRI

In Chapter 4 (Drobinin, Gestel, et al., 2020), I reported the reliability of nine MRI-derived measures of cortical and white matter morphology and integrity. Despite the high prevalence of anxiety and ADHD disorders in our young sample, we found good to excellent reliability for all measures. White matter volume was most consistently reconstructed. Cortical thickness was overall less reliable than surface area. Axial

diffusivity was the least reliable measure experiment-wide. There was regional variability in reconstruction, with some structures being more reliably reconstructed than others. Overall, the take-away is that it is important for researchers to be aware of the underlying reliability of their data, and easily spot if certain results are clustering in areas of low reliability.

One extension to the work would be to explore the reliability of a more comprehensive brain parcellation, specifically the multi modal parcellation devised by Glasser, Coalson, et al., 2016. I suspect the smaller regions of interest would reduce the reliability. On a similar note, reporting the overall reliability of the data, *without* excluding scans that failed QC might have also been of interest to the broader scientific community.

During the writing of this chapter I developed a greater awareness of open science practices and was inspired to open up the data, code, and analysis to the broader scientific community (Drobinin, 2020). While I found the process educational and enjoyable, even individuals who advocate for reproducibility of neuroscience research nevertheless acknowledge that there is a cost of reproducibility (Poldrack, 2019). In other words the field is balancing the incentive structure for taking the time to implement open science practices compared to completing additional projects sooner.

Furthermore, reliability is not the same as generalizability. For example, while cortical thickness estimates might be reliable within-scanner, harmonization techniques might be required if aggregating datasets from across different studies and scanners (Fortin et al., 2018). This is something I encountered in practice in Chapter 5.

Additionally, the relatively lower reliability of fMRI should not be used as the sole basis for making recommendations. Reliability should be considered alongside validity and the research questions at hand (Noble et al., 2019). Others have argued that depending on the measurement, fMRI can be highly reliable (Kragel, Han, Kraynak, Gianaros, & Wager, 2020). To illustrate the point further, a functional connectivity profile, despite being based on connections with 'low reliability', nevertheless offers enough sensitivity to robustly identify individuals with near perfect accuracy (Finn et al., 2015).

Overall, I established that developmental MRI data can be of very high quality and apt for use in developing brain biomarkers. Summaries of reliability are available

for easy lookup by region and hemisphere for any interested researchers.

6.4 Deviation from typical development

In Chapter 5 I accurately and reliably predicted chronological age using neuroanatomical data from a large sample of typically developing youth. The prediction accuracy generalized to an independent cohort of youth enriched for risk of mental illness (FORBOW). Within FORBOW, I found that a number of factors linked to the risks and consequences of mental illness are associated with discrepancies between predicted and chronological age. The spectrum of adverse experiences, from early antecedents for mental illness, to environmental adversity, adolescent onset depression and functional impairment, resulted in an older appearing brain. Overall, the work showed that the brain age gap might be a general indicator of unwellness rather than a marker of specific exposure or outcome.

In order to build better biomarkers, brain models need to be assessed across different samples, research contexts, and populations (Woo et al., 2017). There are a few barriers to this goal. The present study required a sufficient number of developmental scans as well as a phenotype that was consistent across samples. I focused on age because it is present for all data, and as a first foray into machine learning it was beneficial to optimize for an objective marker. However, there would be barriers if the main phenotype of interest was a cognitive test, or a symptom scale, as those can vary significantly between studies. Future work should consider cross-compatibility of measures during the design phase to enhance the ever growing collaborative environment in science.

Finally, the model has been trained and tested on predominantly White samples. Thus when we discuss generalizability we should also consider the built in bias in machine learning algorithms trained on majority groups. Neuroscience, along with other fields, can benefit from improving representation and increased diversity.

6.5 Conclusion

In conclusion, I examined brain correlates of familial risk and early antecedents for mental illness. Furthermore, I established the reliability of widely used MRI measures

in a developmental sample. Finally, I examined how increases in neuroanatomical maturity relate to mental illness. These findings contribute to the field of developmental neuroimaging of at-risk youth.

References

Acheson, A., Wijtenburg, S. A., Rowland, L. M., Winkler, A., Mathias, C. W., Hong, L. E., ... Dougherty, D. M. (2017). Reproducibility of tract-based white matter microstructural measures using the ENIGMA-DTI protocol. *Brain and Behavior*, *7*(2), e00615. doi: 10.1002/brb3.615

Adleman, N. E., Fromm, S. J., Razdan, V., Kayser, R., Dickstein, D. P., Brotman, M. A., ... Leibenluft, E. (2012). Cross-sectional and longitudinal abnormalities in brain structure in children with severe mood dysregulation or bipolar disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(11), 1149–1156. doi: 10.1111/j.1469-7610.2012.02568.x

Adler, C. M., Levine, A. D., DelBello, M. P., & Strakowski, S. M. (2005). Changes in gray matter volume in patients with bipolar disorder. *Biological Psychiatry*, *58*(2), 151–157. doi: 10.1016/j.biopsych.2005.03.022

Al-Chalabi, A., & Lewis, C. M. (2011). Modelling the effects of penetrance and family size on rates of sporadic and familial disease. *Human Heredity*, *71*(4), 281–288. doi: 10.1159/000330167

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., ... Milham, M. P. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, *4*(1), 170181. doi: 10.1038/sdata.2017.181

Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., & Raznahan, A. (2016). Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. *Hum. Brain Mapp.*, *37*(7), 2385–2397.

Alhamud, A., Taylor, P. A., Laughton, B., van der Kouwe, A. J., & Meintjes, E. M. (2015). Motion artifact reduction in pediatric diffusion tensor imaging using fast prospective correction: Motion Artifact Reduction in Pediatric DTI. *Journal of Magnetic Resonance Imaging*, *41*(5), 1353–1364. doi: 10.1002/jmri.24678

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Author.

Armstrong, E., Schleicher, A., Omran, H., Curtis, M., & Zilles, K. (1995 Jan-Feb). The ontogeny of human gyrification. *Cerebral Cortex (New York, N.Y.: 1991)*, *5*(1), 56–63.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. doi: 10.1016/j.tics.2013.12.003

Arseneault, L., Cannon, M., Fisher, H. L., Polanczyk, G., Moffitt, T. E., & Caspi, A. (2011). Childhood trauma and children's emerging psychotic symptoms: A genetically sensitive longitudinal cohort study. *The American Journal of Psychiatry*, *168*(1), 65–72. doi: 10.1176/appi.ajp.2010.10040567

Attell, B. K., Cappelli, C., Manteuffel, B., & Li, H. (2018). Measuring Functional Impairment in Children and Adolescents: Psychometric Properties of the Columbia Impairment Scale (CIS):. *Evaluation & the Health Professions*. doi: 10.1177/0163278718775797

Bagot, K. S., Milin, R., & Kaminer, Y. (2015). Adolescent Initiation of Cannabis Use and Early-Onset Psychosis. *Substance Abuse*, *36*(4), 524–533. doi: 10.1080/08897077.2014.995332

Baiano, M., Perlini, C., Rambaldelli, G., Cerini, R., Dusi, N., Bellani, M., ... Brambilla, P. (2008). Decreased entorhinal cortex volumes in schizophrenia. *Schizophr. Res.*, *102*(1-3), 171–180.

Bartholomeusz, C. F., Whittle, S. L., Montague, A., Ansell, B., McGorry, P. D., Velakoulis, D., ... Wood, S. J. (2013). Sulcogyral patterns and morphological abnormalities of the orbitofrontal cortex in psychosis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *44*, 168–177. doi: 10.1016/j.pnpbp.2013.02.010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bauer, I. E., Sanches, M., Suchting, R., Green, C. E., El Fangary, N. M., Zunta–Soares, G. B., & Soares, J. C. (2014). Amygdala enlargement in unaffected offspring of bipolar parents. *Journal of Psychiatric Research*, *59*, 200–205. doi: 10.1016/j.jpsychires.2014.08.023

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bird, H. R., Shaffer, D., Fisher, P., Gould, M. S., & et al. (1993). The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research*, *3*(3), 167–176.

Blacker, D., & Tsuang, M. T. (1993). Unipolar relatives in bipolar pedigrees: Are they bipolar?. *Psychiatric Genetics*, *3*(1), 5–16.

Bonekamp, D., Nagae, L. M., Degaonkar, M., Matson, M., Abdalla, W. M. A., Barker, P. B., ... Horská, A. (2007). Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences. *Neuroimage*, *34*(2), 733–742.

Bootsman, F., Brouwer, R. M., Schnack, H. G., van Baal, G. C. M., van der Schot, A. C., Vonk, R., ... van Haren, N. E. M. (2015). Genetic and environmental influences on cortical surface area and cortical thickness in bipolar disorder. *Psychological Medicine*, *45*(1), 193–204. doi: 10.1017/S0033291714001251

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive endophenotypes of bipolar disorder: A meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders*, *113*(1-2), 1–20. doi: 10.1016/j.jad.2008.06.009

Borrell, V. (2018). How Cells Fold the Cerebral Cortex. *Journal of Neuroscience*, *38*(4), 776–783. doi: 10.1523/JNEUROSCI.1106-17.2017

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Bremner, J. D., & Vermetten, E. (2001). Stress and development: Behavioral and biological consequences. *Development and Psychopathology*, *13*(3), 473–489. doi: 10.1017/S0954579401003042

Brown, T. T., Kuperman, J. M., Chung, Y., Erhart, M., McCabe, C., Hagler, D. J., ... Dale, A. M. (2012). Neuroanatomical Assessment of Biological Maturity. *Current Biology*, *22*(18), 1693–1698. doi: 10.1016/j.cub.2012.07.002

Bschor, T., Angst, J., Azorin, J. M., Bowden, C. L., Perugi, G., Vieta, E., ... Krüger, S. (2012). Are bipolar disorders underdiagnosed in patients with depressive episodes? Results of the multicenter BRIDGE screening study in Germany. *Journal of Affective Disorders*, *142*(1-3), 45–52. doi: 10.1016/j.jad.2012.03.042

Bubb, E. J., Metzler-Baddeley, C., & Aggleton, J. P. (2018). The cingulum bundle: Anatomy, function, and dysfunction. *Neuroscience & Biobehavioral Reviews*, *92*, 104–127. doi: 10.1016/j.neubiorev.2018.05.008

Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, *23*(2), 724–738. doi: 10.1016/j.neuroimage.2004.06.018

Budde, M. D., Xie, M., Cross, A. H., & Song, S.-K. (2009). Axial Diffusivity Is the Primary Correlate of Axonal Injury in the Experimental Autoimmune Encephalomyelitis Spinal Cord: A Quantitative Pixelwise Analysis. *Journal of Neuroscience*, *29*(9), 2805–2813. doi: 10.1523/JNEUROSCI.4605-08.2009

Cannon, T. D., Chung, Y., He, G., Sun, D., Jacobson, A., van Erp, T. G. M., ... North American Prodrome Longitudinal Study Consortium (2015). Progressive reduction in cortical thickness as psychosis develops: A multisite longitudinal neuroimaging study of youth at elevated clinical risk. *Biological Psychiatry*, *77*(2), 147–157. doi: 10.1016/j.biopsych.2014.05.023

Cao, B., Passos, I. C., Mwangi, B., Bauer, I. E., Zunta-Soares, G. B., Kapczinski, F., & Soares, J. C. (2016). Hippocampal volume and verbal memory performance in late-stage bipolar disorder. *Journal of Psychiatric Research*, *73*, 102–107. doi: 10.1016/j.jpsychires.2015.12.012

Carlson, H. L., Laliberté, C., Brooks, B. L., Hodge, J., Kirton, A., Bello-Espinosa, L., ... Sherman, E. M. (2014). Reliability and variability of diffusion tensor imaging (DTI) tractography in pediatric epilepsy. *Epilepsy & Behavior*, *37*, 116–122. doi: 10.1016/j.yebeh.2014.06.020

Casey, B., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., ... Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. doi: 10.1016/j.dcn.2018.03.001

Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., ... Moffitt, T. E. (2020). Longitudinal Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among Participants in the Dunedin Birth Cohort Study. *JAMA Network Open*, *3*(4), e203221. doi: 10.1001/jamanetworkopen.2020.3221

Caspi, A., Houts, R. M., Belsky, D. W., Harrington, H., Hogan, S., Ramrakha, S., ... Moffitt, T. E. (2016). Childhood forecasting of a small segment of the population with large economic burden. *Nature Human Behaviour*, *1*. doi: 10.1038/s41562-016-0005

Chang, M., Womer, F. Y., Bai, C., Zhou, Q., Wei, S., Jiang, X., ... Wang, F. (2016). Voxel-Based Morphometry in Individuals at Genetic High Risk for Schizophrenia and Patients with Schizophrenia during Their First Episode of Psychosis. *PloS One*, *11*(10), e0163749. doi: 10.1371/journal.pone.0163749

Chen, J., Liu, J., Calhoun, V. D., Arias-Vasquez, A., Zwiers, M. P., Gupta, C. N., ... Turner, J. A. (2014). Exploration of scanning effects in multi-site structural MRI studies. *Journal of Neuroscience Methods*, *230*, 37–50. doi: 10.1016/j.jneumeth.2014.04.023

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi: 10.1145/2939672.2939785

Chung, Y., Addington, J., Bearden, C. E., Cadenhead, K., Cornblatt, B., Mathalon, D. H., ... for the North American Prodrome Longitudinal Study (NAPLS) Consortium and the Pediatric Imaging, Neurocognition, and Genetics (PING) Study Consortium (2018). Use of Machine Learning to Determine Deviance in Neuroanatomical Maturity Associated With Future Psychosis in Youths at Clinically High Risk. *JAMA Psychiatry*, *75*(9), 960. doi: 10.1001/jamapsychiatry.2018.1543

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi: 10.1037/1040-3590.6.4.284

Cole, J. H., & Franke, K. (2017). Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences*, *40*(12), 681–690. doi: 10.1016/j.tins.2017.10.001

Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124. doi: 10.1016/j.neuroimage.2017.07.059

Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., ... Deary, I. J. (2017). Brain age predicts mortality. *Molecular Psychiatry*. doi: 10.1038/mp.2017.62

Collette, F., Van der Linden, M., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2001). The functional anatomy of inhibition processes investigated with the Hayling task. *NeuroImage*, *14*(2), 258–267. doi: 10.1006/nimg.2001.0846

Conus, P., Macneil, C., & McGorry, P. D. (2014). Public health significance of bipolar disorder: Implications for early intervention and prevention. *Bipolar Disorders*, *16*(5), 548–556. doi: 10.1111/bdi.12137

Dager, S. R., Friedman, S. D., Parow, A., Demopulos, C., Stoll, A. L., Lyoo, I. K., ... Renshaw, P. F. (2004). Brain metabolic alterations in medication-free patients with bipolar disorder. *Archives of General Psychiatry*, *61*(5), 450–458. doi: 10.1001/archpsyc.61.5.450

de Lange, A.-M. G., & Cole, J. H. (2020). Commentary: Correction procedures in brain-age prediction. *NeuroImage: Clinical*, 102229. doi: 10.1016/j.nicl.2020.102229

de Nooij, L., Harris, M. A., Hawkins, E. L., Clarke, T.-K., Shen, X., Chan, S. W. Y., ... Whalley, H. C. (2019). Longitudinal trajectories of brain age in young individuals at familial risk of mood disorder. *Wellcome Open Research*, *4*, 206. doi: 10.12688/wellcomeopenres.15617.1

Dean, K., Stevens, H., Mortensen, P. B., Murray, R. M., Walsh, E., & Pedersen, C. B. (2010). Full spectrum of psychiatric outcomes among offspring with parental history of mental disorder. *Archives of General Psychiatry*, *67*(8), 822–829. doi: 10.1001/archgenpsychiatry.2010.86

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Drobinin, V. (2020). *Reliability of youth MRI data, code and analysis.* Zenodo. doi: 10.5281/zenodo.3627320

Drobinin, V., Gestel, H. V., Helmick, C. A., Schmidt, M. H., Bowen, C. V., & Uher, R. (2020). Reliability of multimodal MRI brain measures in youth at risk for mental illness. *Brain and Behavior*, *10*(6), e01609. doi: 10.1002/brb3.1609

Drobinin, V., Slaney, C., Garnham, J., Propper, L., Uher, R., Alda, M., & Hajek, T. (2019). Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders. *Psychological Medicine*, *49*(08), 1308–1315. doi: 10.1017/S0033291718001903

Drobinin, V., Van Gestel, H., Zwicker, A., MacKenzie, L., Cumby, J., Patterson, V. C., ... Uher, R. (2020). Psychotic symptoms are associated with lower cortical folding in youth at risk for mental illness. *Journal of psychiatry & neuroscience: JPN*, *45*(2), 125–133. doi: 10.1503/jpn.180144

Ducharme, S., Albaugh, M. D., Nguyen, T.-V., Hudziak, J. J., Mateos-Pérez, J. M., Labbe, A., ... Karama, S. (2016). Trajectories of cortical thickness maturation in normal brain development — The importance of quality control procedures. *NeuroImage*, *125*(Supplement C), 267–279. doi: 10.1016/j.neuroimage.2015.10.010

Duffy, A., Alda, M., Hajek, T., & Grof, P. (2009). Early course of bipolar disorder in high-risk offspring: Prospective study. *The British Journal of Psychiatry: The Journal of Mental Science*, *195*(5), 457–458. doi: 10.1192/bjp.bp.108.062810

Duffy, A., Alda, M., Hajek, T., Sherry, S. B., & Grof, P. (2010). Early stages in the development of bipolar disorder. *Journal of Affective Disorders*, *121*(1-2), 127–135. doi: 10.1016/j.jad.2009.05.022

Duffy, A., Alda, M., Kutcher, S., Cavazzoni, P., Robertson, C., Grof, E., & Grof, P. (2002). A prospective study of the offspring of bipolar parents responsive and nonresponsive to lithium treatment. *The Journal of Clinical Psychiatry*, *63*(12), 1171–1178.

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., ... Hariri, A. R. (2020). What is the test-retest reliability of common task-fMRI measures? New empirical evidence and a meta-analysis. *bioRxiv*, 681700. doi: 10.1101/681700

Endicott, J., & Spitzer, R. L. (1978). A Diagnostic Interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*, *35*(7), 837–844. doi: 10.1001/archpsyc.1978.01770310043002

Engemann, D. A., Kozynets, O., Sabbagh, D., Lemaître, G., Varoquaux, G., Liem, F., & Gramfort, A. (2020). Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *eLife*, *9*, e54055. doi: 10.7554/eLife.54055

Falkai, P., Honer, W. G., Kamer, T., Dustert, S., Vogeley, K., Schneider-Axmann, T., ... Tepest, R. (2007). Disturbed frontal gyrification within families affected with schizophrenia. *Journal of Psychiatric Research*, *41*(10), 805–813. doi: 10.1016/j.jpsychires.2006.07.018

Favre, P., Pauling, M., Stout, J., Hozer, F., Sarrazin, S., Abé, C., ... Houenou, J. (2019). Widespread white matter microstructural abnormalities in bipolar disorder: Evidence from mega- and meta-analyses across 3033 individuals. *Neuropsychopharmacology*, 1–9. doi: 10.1038/s41386-019-0485-6

Fernández, V., Llinares-Benadero, C., & Borrell, V. (2016). Cerebral cortex expansion and folding: What have we learned? *The EMBO Journal*, *35*(10), 1021–1044. doi: 10.15252/embj.201593701

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. doi: 10.1038/nn.4135

First, M. B. (2015). Structured Clinical Interview for the DSM (SCID). In R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology* (pp. 1–6). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi: 10.1002/9781118625392.wbecp351

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, *97*(20), 11050–11055. doi: 10.1073/pnas.200033797

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, *14*(1), 11–22. doi: 10.1093/cercor/bhg087

Florio, M., & Huttner, W. B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. *Development*, *141*(11), 2182–2194. doi: 10.1242/dev.090571

Foland-Ross, L. C., Bookheimer, S. Y., Lieberman, M. D., Sugar, C. A., Townsend, J. D., Fischer, J., ... Altshuler, L. L. (2012). Normal amygdala activation but deficient ventrolateral prefrontal activation in adults with bipolar disorder during euthymia. *NeuroImage*, *59*(1), 738–744. doi: 10.1016/j.neuroimage.2011.07.054

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327. doi: 10.1016/j.neuroimage.2010.07.033

Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., ... Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, *167*, 104–120. doi: 10.1016/j.neuroimage.2017.11.024

Frangou, S. (2011). Brain structural and functional correlates of resilience to Bipolar Disorder. *Frontiers in Human Neuroscience*, *5*, 184. doi: 10.3389/fnhum.2011.00184

Franke, K., & Gaser, C. (2019). Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Frontiers in Neurology*, *10*. doi: 10.3389/fneur.2019.00789

Franke, K., Luders, E., May, A., Wilke, M., & Gaser, C. (2012). Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *NeuroImage*, *63*(3), 1305–1312. doi: 10.1016/j.neuroimage.2012.08.001

Friston, K., Brown, H. R., Siemerkus, J., & Stephan, K. E. (2016). The dysconnection hypothesis (2016). *Schizophrenia Research*, *176*(2-3), 83–94. doi: 10.1016/j.schres.2016.07.014

Fung, G., Deng, Y., Zhao, Q., Li, Z., Qu, M., Li, K., ... Chan, R. C. K. (2015). Distinguishing bipolar and major depressive disorders by brain structural morphometry: A pilot study. *BMC Psychiatry*, *15*, 298. doi: 10.1186/s12888-015-0685-5

Fusar-Poli, P., Bonoldi, I., Yung, A. R., Borgwardt, S., Kempton, M. J., Valmaggia, L., ... McGuire, P. (2012). Predicting Psychosis: Meta-analysis of Transition Outcomes in Individuals at High Clinical Risk. *Archives of General Psychiatry*, *69*(3), 220–229. doi: 10.1001/archgenpsychiatry.2011.1472

Fux, L., Walger, P., Schimmelmann, B. G., & Schultze-Lutter, F. (2013). The Schizophrenia Proneness Instrument, Child and Youth version (SPI-CY): Practicability and discriminative validity. *Schizophrenia Research*, *146*(1-3), 69–78. doi: 10.1016/j.schres.2013.02.014

Gershon, E. S., Hamovit, J., Guroff, J. J., Dibble, E., Leckman, J. F., Sceery, W., ... Bunney, W. E. (1982). A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. *Archives of General Psychiatry*, *39*(10), 1157–1167.

Gerson, A. C., Gerring, J. P., Freund, L., Joshi, P. T., Capozzoli, J., Brady, K., & Denckla, M. B. (1996). The Children's Affective Lability Scale: A psychometric evaluation of reliability. *Psychiatry Research*, *65*(3), 189–198. doi: 10.1016/s0165-1781(96)02851-x

Ghaemi, S. N., Sachs, G. S., Chiou, A. M., Pandurangi, A. K., & Goodwin, K. (1999 Jan-Mar). Is bipolar disorder still underdiagnosed? Are antidepressants overutilized? *Journal of Affective Disorders*, *52*(1-3), 135–144.

Gianaros, P. J., Kuan, D. C.-H., Marsland, A. L., Sheu, L. K., Hackman, D. A., Miller, K. G., & Manuck, S. B. (2017). Community Socioeconomic Disadvantage in Midlife Relates to Cortical Morphology via Neuroendocrine and Cardiometabolic Pathways. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(1), 460–473. doi: 10.1093/cercor/bhv233

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... 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., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., ... Van Essen, D. C. (2016). The Human Connectome Project's neuroimaging approach. *Nature Neuroscience*, *19*(9), 1175–1187. doi: 10.1038/nn.4361

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. doi: 10.1016/j.neuroimage.2013.04.127

Global Burden of Disease Study 2013 Collaborators. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)*, *386*(9995), 743–800. doi: 10.1016/S0140-6736(15)60692-4

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., ... Etkin, A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA psychiatry*, *72*(4), 305–315. doi: 10.1001/jamapsychiatry.2014.2206

Gottesman, I. I., & Gould, T. D. (2003). The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *American Journal of Psychiatry*, *160*(4), 636–645. doi: 10.1176/appi.ajp.160.4.636

Guo, S., Iwabuchi, S., Balain, V., Feng, J., Liddle, P., & Palaniyappan, L. (2015). Cortical folding and the potential for prognostic neuroimaging in schizophrenia. *British Journal of Psychiatry*, *207*(5), 458–459. doi: 10.1192/bjp.bp.114.155796

Haijma, S. V., Van Haren, N., Cahn, W., Koolschijn, P. C. M. P., Hulshoff Pol, H. E., & Kahn, R. S. (2013). Brain volumes in schizophrenia: A meta-analysis in over 18 000 subjects. *Schizophrenia Bulletin*, *39*(5), 1129–1138. doi: 10.1093/schbul/sbs118

Hajek, T., Alda, M., Hajek, E., & Ivanoff, J. (2013). Functional neuroanatomy of response inhibition in bipolar disorders–combined voxel based and cognitive performance meta-analysis. *Journal of Psychiatric Research*, *47*(12), 1955–1966. doi: 10.1016/j.jpsychires.2013.08.015

Hajek, T., Bauer, M., Simhandl, C., Rybakowski, J., O'Donovan, C., Pfennig, A., ... Alda, M. (2014). Neuroprotective effect of lithium on hippocampal volumes in bipolar disorder independent of long-term treatment response. *Psychological Medicine*, *44*(3), 507–517. doi: 10.1017/S0033291713001165

Hajek, T., Bernier, D., Slaney, C., Propper, L., Schmidt, M., Carrey, N., ... Alda, M. (2008). A comparison of affected and unaffected relatives of patients with bipolar disorder using proton magnetic resonance spectroscopy. *Journal of Psychiatry & Neuroscience*, *33*(6), 531–540.

Hajek, T., Calkin, C., Blagdon, R., Slaney, C., & Alda, M. (2015). Type 2 diabetes mellitus: A potentially modifiable risk factor for neurochemical brain changes in bipolar disorders. *Biological Psychiatry*, *77*(3), 295–303. doi: 10.1016/j.biopsych.2013.11.007

Hajek, T., Cooke, C., Kopecek, M., Novak, T., Hoschl, C., & Alda, M. (2015). Using structural MRI to identify individuals at genetic risk for bipolar disorders: A 2-cohort, machine learning study. *Journal of psychiatry & neuroscience: JPN*, *40*(5), 316–324.

Hajek, T., Cullis, J., Novak, T., Kopecek, M., Blagdon, R., Propper, L., ... Alda, M. (2013). Brain structural signature of familial predisposition for bipolar disorder: Replicable evidence for involvement of the right inferior frontal gyrus. *Biological Psychiatry*, *73*(2), 144–152. doi: 10.1016/j.biopsych.2012.06.015

Hajek, T., Cullis, J., Novak, T., Kopecek, M., Höschl, C., Blagdon, R., ... Alda, M. (2012). Hippocampal volumes in bipolar disorders: Opposing effects of illness burden and lithium treatment. *Bipolar Disorders*, *14*(3), 261–270. doi: 10.1111/j.1399-5618.2012.01013.x

Hajek, T., & Weiner, M. W. (2016). Neuroprotective Effects of Lithium in Human Brain? Food for Thought. *Current Alzheimer Research*, *13*(8), 862–872.

Hamakawa, H., Kato, T., Murashita, J., & Kato, N. (1998). Quantitative proton magnetic resonance spectroscopy of the basal ganglia in patients with affective disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *248*(1), 53–58. doi: 10.1007/s004060050017

Han, L. K. M., Dinga, R., Hahn, T., Ching, C. R. K., Eyler, L. T., Aftanas, L., ... Schmaal, L. (2020). Brain aging in major depressive disorder: Results from the ENIGMA major depressive disorder working group. *Molecular Psychiatry*, 1–16. doi: 10.1038/s41380-020-0754-0

Hart, H., & Rubia, K. (2012). Neuroimaging of child abuse: A critical review. *Frontiers in Human Neuroscience*, *6*. doi: 10.3389/fnhum.2012.00052

Herting, M. M., Gautam, P., Spielberg, J. M., Dahl, R. E., & Sowell, E. R. (2015). A longitudinal study: Changes in cortical thickness and surface area during pubertal maturation. *PLoS ONE*, *10*(3), 1–14. doi: 10.1371/journal.pone.0119774

Herting, M. M., Johnson, C., Mills, K. L., Vijayakumar, N., Dennison, M., Liu, C., ... Tamnes, C. K. (2018). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *NeuroImage*, *172*, 194–205. doi: 10.1016/j.neuroimage.2018.01.020

Hibar, D. P., Adams, H. H. H., Jahanshad, N., Chauhan, G., Stein, J. L., Hofer, E., ... Ikram, M. A. (2017). Novel genetic loci associated with hippocampal volume. *Nature Communications*, *8*, 13624. doi: 10.1038/ncomms13624

Hibar, D. P., Westlye, L. T., Doan, N. T., Jahanshad, N., Cheung, J. W., Ching, C. R. K., ... Andreassen, O. A. (2018). Cortical abnormalities in bipolar disorder: An MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular Psychiatry*, *23*(4), 932–942. doi: 10.1038/mp.2017.73

Hibar, D. P., Westlye, L. T., van Erp, T. G. M., Rasmussen, J., Leonardo, C. D., Faskowitz, J., ... Andreassen, O. A. (2016). Subcortical volumetric abnormalities in bipolar disorder. *Molecular Psychiatry*, *21*(12), 1710–1716. doi: 10.1038/mp.2015.227

Hillegers, M. H., Reichart, C. G., Wals, M., Verhulst, F. C., Ormel, J., & Nolen, W. A. (2005). Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. *Bipolar Disorders*, *7*(4), 344–350. doi: 10.1111/j.1399-5618.2005.00215.x

Hogstrom, L. J., Westlye, L. T., Walhovd, K. B., & Fjell, A. M. (2013). The Structure of the Cerebral Cortex Across Adult Life: Age-Related Patterns of Surface Area, Thickness, and Gyrification. *Cerebral Cortex*, *23*(11), 2521–2530. doi: 10.1093/cercor/bhs231

Homayoun, H., & Moghaddam, B. (2008). Orbitofrontal cortex neurons as a common target for classic and glutamatergic antipsychotic drugs. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(46), 18041–18046. doi: 10.1073/pnas.0806669105

Hong, Y.-W., Yoo, Y., Han, J., Wager, T. D., & Woo, C.-W. (2019). False-positive neuroimaging: Undisclosed flexibility in testing spatial hypotheses allows presenting anything as a replicated finding. *NeuroImage*, *195*, 384–395. doi: 10.1016/j.neuroimage.2019.03.070

Huttenlocher, P. R. (1994). Synaptogenesis in human cerebral cortex. In *Human behavior and the developing brain* (pp. 137–152). New York, NY, US: The Guilford Press.

Im, K., Lee, J.-M., Lyttelton, O., Kim, S. H., Evans, A. C., & Kim, S. I. (2008). Brain Size and Cortical Structure in the Adult Human Brain. *Cerebral Cortex*, *18*(9), 2181–2191. doi: 10.1093/cercor/bhm244

Iscan, Z., Jin, T. B., Kendrick, A., Szeglin, B., Lu, H., Trivedi, M., ... DeLorenzo, C. (2015). Test–retest reliability of freesurfer measurements within and between sites: Effects of visual approval process. *Human Brain Mapping*, *36*(9), 3472–3485. doi: 10.1002/hbm.22856

Isomura, S., Hashimoto, R., Nakamura, M., Hirano, Y., Yamashita, F., Jimbo, S., ... Onitsuka, T. (2017). Altered sulcogyral patterns of orbitofrontal cortex in a large cohort of patients with schizophrenia. *npj Schizophrenia*, *3*(1), 3. doi: 10.1038/s41537-016-0008-y

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY: Springer New York. doi: 10 .1007/978-1-4614-7138-7

Janssen, J., Alemán-Gómez, Y., Schnack, H., Balaban, E., Pina-Camacho, L., Alfaro-Almagro, F., ... Desco, M. (2014). Cortical morphology of adolescents with bipolar disorder and with schizophrenia. *Schizophrenia Research*, *158*(1-3), 91–99. doi: 10.1016/j.schres.2014.06.040

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. doi: 10.1016/j.neuroimage.2011.09 .015

Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Maser, J., Solomon, D. A., ... Keller, M. B. (2002). The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Archives of General Psychiatry*, *59*(6), 530–537.

Judd, L. L., Schettler, P. J., Solomon, D. A., Maser, J. D., Coryell, W., Endicott, J., & Akiskal, H. S. (2008). Psychosocial disability and work role function compared across the long-term course of bipolar I, bipolar II and unipolar major depressive disorders. *Journal of Affective Disorders*, *108*(1-2), 49–58. doi: 10.1016/j.jad.2007.06.014

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., ... Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(7), 980–988. doi: 10.1097/00004583-199707000-00021

Kaufmann, T., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I., ... Westlye, L. T. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, *22*(10), 1617–1623. doi: 10.1038/s41593-019-0471-7

Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., ... Donohoe, G. (2017). Widespread white matter microstructural differences in schizophrenia across 4322 individuals: Results from the ENIGMA Schizophrenia DTI Working Group. *Molecular Psychiatry*. doi: 10.1038/mp.2017.170

Kempton, M. J., Haldane, M., Jogia, J., Grasby, P. M., Collier, D., & Frangou, S. (2009). Dissociable Brain Structural Changes Associated with Predisposition, Resilience, and Disease Expression in Bipolar Disorder. *Journal of Neuroscience*, *29*(35), 10863–10868. doi: 10.1523/JNEUROSCI.2204-09.2009

Kempton, M. J., Salvador, Z., Munafò, M. R., Geddes, J. R., Simmons, A., Frangou, S., & Williams, S. C. R. (2011). Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Archives of General Psychiatry*, *68*(7), 675–690. doi: 10.1001/archgenpsychiatry.2011.60

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593. doi: 10.1001/archpsyc.62.6.593

Kim-Cohen, J., Caspi, A., Moffitt, T. E., Harrington, H., Milne, B. J., & Poulton, R. (2003). Prior Juvenile Diagnoses in Adults With Mental Disorder: Developmental Follow-Back of a Prospective-Longitudinal Cohort. *Archives of General Psychiatry*, *60*(7), 709. doi: 10.1001/archpsyc.60.7.709

Klapwijk, E. T., van de Kamp, F., van der Meulen, M., Peters, S., & Wierenga, L. M. (2019). Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *NeuroImage*, *189*, 116–129. doi: 10.1016/j.neuroimage.2019.01.014

Klosterkötter, J., Hellmich, M., Steinmeyer, E. M., & Schultze-Lutter, F. (2001). Diagnosing schizophrenia in the initial prodromal phase. *Archives of General Psychiatry*, *58*(2), 158–164.

Klyachko, V. A., & Stevens, C. F. (2003). Connectivity optimization and the positioning of cortical areas. *Proceedings of the National Academy of Sciences*, *100*(13), 7937–7941. doi: 10.1073/pnas.0932745100

Konarski, J. Z., McIntyre, R. S., Kennedy, S. H., Rafi-Tari, S., Soczynska, J. K., & Ketter, T. A. (2008). Volumetric neuroimaging investigations in mood disorders: Bipolar disorder versus major depressive disorder. *Bipolar Disorders*, *10*(1), 1–37. doi: 10.1111/j.1399-5618.2008.00435.x

Kong, L., Herold, C. J., Lässer, M. M., Schmid, L. A., Hirjak, D., Thomann, P. A., ... Schröder, J. (2015). Association of Cortical Thickness and Neurological Soft Signs in Patients with Chronic Schizophrenia and Healthy Controls. *Neuropsychobiology*, *71*(4), 225–233. doi: 10.1159/000382020

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. doi: 10.1016/j.jcm.2016.02.012

Kragel, P. A., Han, X., Kraynak, T., Gianaros, P. J., & Wager, T. D. (2020). *fMRI can be highly reliable, but it depends on what you measure* (Preprint). PsyArXiv.

Kroenke, C. D., & Bayly, P. V. (2018). How Forces Fold the Cerebral Cortex. *Journal of Neuroscience*, *38*(4), 767–775. doi: 10.1523/JNEUROSCI.1105-17.2017

Kuhn, M., & Wickham, H. (2020). *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.* [Manual].

Kulynych, J. J., Luevano, L. F., Jones, D. W., & Weinberger, D. R. (1997). Cortical abnormality in schizophrenia: An in vivo application of the gyrification index. *Biological Psychiatry*, *41*(10), 995–999. doi: 10.1016/S0006-3223(96)00292-2

Ladouceur, C. D., Almeida, J. R. C., Birmaher, B., Axelson, D. A., Nau, S., Kalas, C., ... Phillips, M. L. (2008). Subcortical gray matter volume abnormalities in healthy bipolar offspring: Potential neuroanatomical risk marker for bipolar disorder? *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(5), 532–539. doi: 10.1097/CHI.0b013e318167656e

Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., Paulus, M. P., & Tulsa 1000 Investigators. (2018). A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers in Aging Neuroscience*, *10*, 317. doi: 10.3389/fnagi.2018.00317

Lewis, J. D., Evans, A. C., & Tohka, J. (2018). T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance. *NeuroImage*, *173*, 341–350. doi: 10.1016/j.neuroimage.2018.02.050

Lin, A., Ching, C. R. K., Vajdi, A., Sun, D., Jonas, R. K., Jalbrzikowski, M., ... Bearden, C. E. (2017). Mapping 22q11.2 Gene Dosage Effects on Brain Morphometry. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(26), 6183–6199. doi: 10.1523/JNEUROSCI.3759-16.2017

Lin, K., Xu, G., Wong, N. M., Wu, H., Li, T., Lu, W., ... Lee, T. M. (2015). A Multi-Dimensional and Integrative Approach to Examining the High-Risk and Ultra-High-Risk Stages of Bipolar Disorder. *EBioMedicine*, *2*(8), 919–928. doi: 10.1016/j.ebiom.2015.06.027

Liu, B., Zhang, X., Cui, Y., Qin, W., Tao, Y., Li, J., ... Jiang, T. (2017). Polygenic Risk for Schizophrenia Influences Cortical Gyrification in 2 Independent General Populations. *Schizophrenia Bulletin*, *43*(3), 673–680. doi: 10.1093/schbul/sbw051

Lopez de Lara, C., Jaitovich-Groisman, I., Cruceanu, C., Mamdani, F., Lebel, V., Yerko, V., ... Turecki, G. (2010). Implication of synapse-related genes in bipolar disorder by linkage and gene expression analyses. *The International Journal of Neuropsychopharmacology*, *13*(10), 1397–1410. doi: 10.1017/S1461145710000714

MacKenzie, L. E., Abidi, S., Fisher, H. L., Propper, L., Bagnell, A., Morash-Conway, J., ... Uher, R. (2016). Stimulant Medication and Psychotic Symptoms in Offspring of Parents With Mental Illness. *Pediatrics*, *137*(1), e20152486. doi: 10.1542/peds.2015-2486

MacKenzie, L. E., Patterson, V. C., Zwicker, A., Drobinin, V., Fisher, H. L., Abidi, S., ... Uher, R. (2017). Hot and cold executive functions in youth with psychotic symptoms. *Psychological Medicine*, *47*(16), 2844–2853. doi: 10.1017/S0033291717001374

Madan, C. R., & Kensinger, E. A. (2017). Test–retest reliability of brain morphology estimates. *Brain Informatics*, *4*(2), 107–121. doi: 10.1007/s40708-016-0060-4

Marsh, R., Zhu, H., Schultz, R. T., Quackenbush, G., Royal, J., Skudlarski, P., & Peterson, B. S. (2006). A developmental fMRI study of self-regulatory control. *Human Brain Mapping*, *27*(11), 848–863. doi: 10.1002/hbm.20225

Matsuo, K., Kopecek, M., Nicoletti, M. A., Hatch, J. P., Watanabe, Y., Nery, F. G., ... Soares, J. C. (2012). New structural brain imaging endophenotype in bipolar disorder. *Molecular Psychiatry*, *17*(4), 412–420. doi: 10.1038/mp.2011.3

McDermott, C. L., Seidlitz, J., Nadig, A., Liu, S., Clasen, L. S., Blumenthal, J. D., ... Raznahan, A. (2019). Longitudinally Mapping Childhood Socioeconomic Status Associations with Cortical and Subcortical Morphology. *The Journal of Neuroscience*, *39*(8), 1365–1373. doi: 10.1523/JNEUROSCI.1808-18.2018

McGrath, B. M., Wessels, P. H., Bell, E. C., Ulrich, M., & Silverstone, P. H. (2004). Neurobiological Findings in Bipolar II Disorder Compared with Findings in Bipolar I Disorder. *The Canadian Journal of Psychiatry*, *49*(12), 794–801. doi: 10.1177/070674370404901202

McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., & Cardno, A. (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of General Psychiatry*, *60*(5), 497–502. doi: 10.1001/archpsyc.60.5.497

McIntosh, A. M., Moorhead, T. W. J., McKirdy, J., Hall, J., Sussmann, J. E. D., Stanfield, A. C., ... Lawrie, S. M. (2009). Prefrontal gyral folding and its cognitive correlates in bipolar disorder and schizophrenia. *Acta Psychiatrica Scandinavica*, *119*(3), 192–198. doi: 10.1111/j.1600-0447.2008.01286.x

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, *12*(3), 131–143.

Miller, T. J., McGlashan, T. H., Rosen, J. L., Cadenhead, K., Cannon, T., Ventura, J., ... Woods, S. W. (2003). Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: Predictive validity, interrater reliability, and training to reliability. *Schizophrenia Bulletin*, *29*(4), 703–715.

Mills, K. L., Goddings, A.-L., Herting, M. M., Meuwese, R., Blakemore, S.-J., Crone, E. A., ... Tamnes, C. K. (2016). Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *NeuroImage*, *141*, 273–281. doi: 10.1016/j.neuroimage.2016.07.044

Mori, S., Wakana, S., van Zijl, P. C. M., & Nagae-Poetscher, L. M. (2005). *MRI Atlas of Human White Matter*. Elsevier.

Mowinckel, A. M., & Vidal-Piñeiro, D. (2019). Visualisation of Brain Statistics with R-packages ggseg and ggseg3d. *arXiv:1912.08200 [stat]*.

Moylan, S., Maes, M., Wray, N. R., & Berk, M. (2013). The neuroprogressive nature of major depressive disorder: Pathways to disease evolution and resistance, and therapeutic implications. *Molecular Psychiatry*, *18*(5), 595–606. doi: 10.1038/mp.2012.33

Nanda, P., Tandon, N., Mathew, I. T., Giakoumatos, C. I., Abhishekh, H. A., Clementz, B. A., ... Keshavan, M. S. (2014). Local Gyrification Index in Probands with Psychotic Disorders and Their First-Degree Relatives. *Biological Psychiatry*, *76*(6), 447–455. doi: 10.1016/j.biopsych.2013.11.018

Nenadic, I., Maitra, R., Dietzek, M., Langbein, K., Smesny, S., Sauer, H., & Gaser, C. (2015). Prefrontal gyrification in psychotic bipolar I disorder vs. schizophrenia. *Journal of Affective Disorders*, *185*, 104–107. doi: 10.1016/j.jad.2015.06.014

Nesvåg, R., Schaer, M., Haukvik, U. K., Westlye, L. T., Rimol, L. M., Lange, E. H., ... Eliez, S. (2014). Reduced brain cortical folding in schizophrenia revealed in two independent samples. *Schizophrenia research*, *152*(2-3), 333–8. doi: 10.1016/j.schres.2013.11.032

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., ... Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*(3), 299–303. doi: 10.1038/nn.4500

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, *203*, 116157. doi: 10.1016/j.neuroimage.2019.116157

Nunes, A., Schnack, H. G., Ching, C. R. K., Agartz, I., Akudjedu, T. N., Alda, M., ... Hajek, T. (2018). Using structural MRI to identify bipolar disorders – 13 site machine learning study in 3020 individuals from the ENIGMA Bipolar Disorders Working Group. *Molecular Psychiatry*, 1–14. doi: 10.1038/s41380-018-0228-9

Opel, N., Goltermann, J., Hermesdorf, M., Berger, K., Baune, B. T., & Dannlowski, U. (2020). Cross-Disorder Analysis of Brain Structural Abnormalities in Six Major Psychiatric Disorders – A Secondary Analysis of Mega- and Meta-Analytical Findings from the ENIGMA Consortium. *Biological Psychiatry*. doi: 10.1016/j.biopsych.2020.04.027

Ortiz, A., Bradler, K., Slaney, C., Garnham, J., Ruzickova, M., O'Donovan, C., ... Alda, M. (2011). An admixture analysis of the age at index episodes in bipolar disorder. *Psychiatry Research*, *188*(1), 34–39. doi: 10.1016/j.psychres.2010.10.033

Paccalet, T., Gilbert, E., Berthelot, N., Marquet, P., Jomphe, V., Lussier, D., ... Maziade, M. (2016). Liability indicators aggregate many years before transition to illness in offspring descending from kindreds affected by schizophrenia or bipolar disorder. *Schizophrenia Research*, *175*(1), 186–192. doi: 10.1016/j.schres.2016.04.038

Palaniyappan, L., & Liddle, P. F. (2014). Diagnostic Discontinuity in Psychosis: A Combined Study of Cortical Gyrification and Functional Connectivity. *Schizophrenia Bulletin*, *40*(3), 675–684. doi: 10.1093/schbul/sbt050

Palaniyappan, L., Mallikarjun, P., Joseph, V., White, T. P., & Liddle, P. F. (2011). Folding of the Prefrontal Cortex in Schizophrenia: Regional Differences in Gyrification. *Biological Psychiatry*, *69*(10), 974–979. doi: 10.1016/j.biopsych.2010.12.012

Palaniyappan, L., Marques, T. R., Taylor, H., Handley, R., Mondelli, V., Bonaccorso, S., ... Dazzan, P. (2013). Cortical Folding Defects as Markers of Poor Treatment Response in First-Episode Psychosis. *Jama Psychiatry*, *70*(10), 1031–1040. doi: 10.1001/jamapsychiatry.2013.203

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., ... Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex*, *19*(11), 2728–35. doi: 10.1093/cercor/bhp026

Pantelis, C., Velakoulis, D., McGorry, P. D., Wood, S. J., Suckling, J., Phillips, L. J., ... McGuire, P. K. (2003). Neuroanatomical abnormalities before and after onset of psychosis: A cross-sectional and longitudinal MRI comparison. *Lancet*, *361*(9354), 281–288. doi: 10.1016/S0140-6736(03)12323-9

Patterson, V. C., Pencer, A., Pavlova, B., Awadia, A., MacKenzie, L. E., Zwicker, A., ... Uher, R. (2020). Youth Experience Tracker Instrument: A self-report measure of developmental antecedents to severe mental illness. *Early Intervention in Psychiatry*, *n/a*(n/a). doi: 10.1111/eip.13007

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. doi: 10.1038/nrn2513

Pavlova, B., Perlis, R. H., Alda, M., & Uher, R. (2015). Lifetime prevalence of anxiety disorders in people with bipolar disorder: A systematic review and meta-analysis. *The Lancet. Psychiatry*, *2*(8), 710–717. doi: 10.1016/S2215-0366(15)00112-1

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., & Smith, S. M. (2019). Accurate brain age prediction with lightweight deep neural networks. *bioRxiv*, 2019.12.17.879346. doi: 10.1101/2019.12.17.879346

Polanczyk, G., Moffitt, T. E., Arseneault, L., Cannon, M., Ambler, A., Keefe, R. S. E., ... Caspi, A. (2010). Etiological and clinical features of childhood psychotic symptoms: Results from a birth cohort. *Archives of General Psychiatry*, *67*(4), 328–338. doi: 10.1001/archgenpsychiatry.2010.14

Poldrack, R. A. (2019). The Costs of Reproducibility. *Neuron*, *101*(1), 11–14.

Poulton, R., Caspi, A., Moffitt, T. E., Cannon, M., Murray, R., & Harrington, H. (2000). Children's self-reported psychotic symptoms and adult schizophreniform disorder: A 15-year longitudinal study. *Archives of General Psychiatry*, *57*(11), 1053–1058.

Radua, J., Ramella-Cravaro, V., Ioannidis, J. P., Reichenberg, A., Phiphopthatsanee, N., Amir, T., ... Fusar-Poli, P. (2018). What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry*, *17*(1), 49–66. doi: 10.1002/wps.20490

Rakic, P., Ayoub, A. E., Breunig, J. J., & Dominguez, M. H. (2009). Decision by division: Making cortical maps. *Trends in Neurosciences*, *32*(5), 291–301. doi: 10.1016/j.tins.2009.01.007

Rapoport, J. L., Giedd, J. N., & Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: Update 2012. *Molecular Psychiatry*, *17*(12), 1228–1238. doi: 10.1038/mp.2012.23

Rasic, D., Hajek, T., Alda, M., & Uher, R. (2014). Risk of Mental Illness in Offspring of Parents With Schizophrenia, Bipolar Disorder, and Major Depressive Disorder: A Meta-Analysis of Family High-Risk Studies. *Schizophrenia Bulletin*, *40*(1), 28–38. doi: 10.1093/schbul/sbt114

Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J., & 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

Roberts, G., Green, M. J., Breakspear, M., McCormack, C., Frankland, A., Wright, A., ... Mitchell, P. B. (2013). Reduced inferior frontal gyrus activation during response inhibition to emotional stimuli in youth at high risk of bipolar disorder. *Biological Psychiatry*, *74*(1), 55–61. doi: 10.1016/j.biopsych.2012.11.004

Roberts, G., Lenroot, R., Frankland, A., Yeung, P. K., Gale, N., Wright, A., ... Mitchell, P. B. (2016). Abnormalities in left inferior frontal gyral thickness and parahippocampal gyral volume in young people at high genetic risk for bipolar disorder. *Psychological Medicine*, *46*(10), 2083–2096. doi: 10.1017/S0033291716000507

Roberts, G., Lord, A., Frankland, A., Wright, A., Lau, P., Levy, F., ... Breakspear, M. (2017). Functional Dysconnection of the Inferior Frontal Gyrus in Young People With Bipolar Disorder or at Genetic High Risk. *Biological Psychiatry*, *81*(8), 718–727. doi: 10.1016/j.biopsych.2016.08.018

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in Neurobiology*, *86*(3), 216–244. doi: 10.1016/j.pneurobio.2008.09.001

Romer, A. L., Elliott, M. L., Knodt, A. R., Sison, M. L., Ireland, D., Houts, R., ... Hariri, A. R. (2020). Pervasively Thinner Neocortex as a Transdiagnostic Feature of General Psychopathology. *The American Journal of Psychiatry*, appiajp202019090934. doi: 10.1176/appi.ajp.2020.19090934

RStudio Team. (2019). *RStudio: Integrated development environment for R.* Boston, MA.

Sarıçiçek, A., Yalın, N., Hıdıroğlu, C., Çavuşoğlu, B., Taş, C., Ceylan, D., ... Özerdem, A. (2015). Neuroanatomical correlates of genetic risk for bipolar disorder: A voxel-based morphometry study in bipolar type I patients and healthy first degree relatives. *Journal of Affective Disorders*, *186*, 110–118. doi: 10.1016/j.jad.2015.06.055

Sasabayashi, D., Takayanagi, Y., Takahashi, T., Koike, S., Yamasue, H., Katagiri, N., ... Suzuki, M. (2017). Increased Occipital Gyrification and Development of Psychotic Disorders in Individuals With an At-Risk Mental State: A Multicenter Study. *Biological Psychiatry*, *82*(10), 737–745. doi: 10.1016/j.biopsych.2017.05.018

Schaer, M., Cuadra, M., Tamarit, L., Lazeyras, F., Eliez, S., & Thiran, J.-P. (2008). A Surface-Based Approach to Quantify Local Cortical Gyrification. *IEEE Transactions on Medical Imaging*, *27*(2), 161–170. doi: 10.1109/TMI.2007.903576

Schaer, M., Debbane, M., Cuadra, M. B., Ottet, M.-C., Glaser, B., Thiran, J.-P., & Eliez, S. (2009). Deviant trajectories of cortical maturation in 22q11.2 deletion syndrome (22q11DS): A cross-sectional and longitudinal study. *Schizophrenia Research*, *115*(2-3), 182–190. doi: 10.1016/j.schres.2009.09.016

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., ... Veltman, D. J. (2017). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Molecular Psychiatry*, *22*(6), 900–909. doi: 10.1038/mp.2016.60

Schmaal, L., Veltman, D. J., van Erp, T. G. M., Sämann, P. G., Frodl, T., Jahanshad, N., ... Hibar, D. P. (2016). Subcortical brain alterations in major depressive disorder: Findings from the ENIGMA Major Depressive Disorder working group. *Molecular Psychiatry*, *21*(6), 806–812. doi: 10.1038/mp.2015.69

Schnack, H. G., van Haren, N. E., Nieuwenhuis, M., Hulshoff Pol, H. E., Cahn, W., & Kahn, R. S. (2016). Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. *American Journal of Psychiatry*, *173*(6), 607–616. doi: 10.1176/appi.ajp.2015.15070922

Schultz, C. C., Wagner, G., Koch, K., Gaser, C., Roebel, M., Schachtzabel, C., ... Schloesser, R. G. M. (2013). The visual cortex in schizophrenia: Alterations of gyrification rather than cortical thickness-a combined cortical shape analysis. *Brain Structure & Function*, *218*(1), 51–58. doi: 10.1007/s00429-011-0374-1

Schultze-Lutter, F. (2009). Subjective symptoms of schizophrenia in research and the clinic: The basic symptom concept. *Schizophrenia Bulletin*, *35*(1), 5–8. doi: 10.1093/schbul/sbn139

Schultze-Lutter, F., Ruhrmann, S., Fusar-Poli, P., Bechdolf, A., Schimmelmann, B. G., & Klosterkötter, J. (2012). Basic symptoms and the prediction of first-episode psychosis. *Current Pharmaceutical Design*, *18*(4), 351–357.

Selemon, L. D., Kleinman, J. E., Herman, M. M., & Goldman-Rakic, P. S. (2002). Smaller frontal gray matter volume in postmortem schizophrenic brains. *The American Journal of Psychiatry*, *159*(12), 1983–1991. doi: 10.1176/appi.ajp.159.12.1983

Selemon, L. D., & Zecevic, N. (2015). Schizophrenia: A tale of two critical periods for prefrontal cortical development. *Translational Psychiatry*, *5*(8), e623. doi: 10.1038/tp.2015.115

Selvaraj, S., Arnone, D., Job, D., Stanfield, A., Farrow, T. F., Nugent, A. C., ... McIntosh, A. M. (2012). Grey matter differences in bipolar disorder: A meta-analysis of voxel-based morphometry studies. *Bipolar Disorders*, *14*(2), 135–145. doi: 10.1111/j.1399-5618.2012.01000.x

Selzam, S., Coleman, J. R. I., Caspi, A., Moffitt, T. E., & Plomin, R. (2018). A polygenic p factor for major psychiatric disorders. *Translational Psychiatry*, *8*(1), 205. doi: 10.1038/s41398-018-0217-4

Seo, H., & Lee, D. (2012). Neural basis of learning and preference during social decision-making. *Current Opinion in Neurobiology*, *22*(6), 990–995. doi: 10.1016/j.conb.2012.05.010

Shollenbarger, S. G., Price, J., Wieser, J., & Lisdahl, K. (2015). Impact of cannabis use on prefrontal and parietal cortex gyrification and surface area in adolescents and emerging adults. *Developmental Cognitive Neuroscience*, *16*, 46–53. doi: 10.1016/j.dcn.2015.07.004

Silverstone, P. H., Asghar, S. J., O'Donnell, T., Ulrich, M., & Hanstock, C. C. (2004). Lithium and valproate protect against dextro-amphetamine induced brain choline concentration changes in bipolar disorder patients. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, *5*(1), 38–44.

Singh, M. K., Kelley, R. G., Howe, M. E., Reiss, A. L., Gotlib, I. H., & Chang, K. D. (2014). Reward processing in healthy offspring of parents with bipolar disorder. *JAMA psychiatry*, *71*(10), 1148–1156. doi: 10.1001/jamapsychiatry.2014.1031

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E., & Miller, K. L. (2019). Estimation of brain age delta from brain imaging. *NeuroImage*, *200*, 528–539. doi: 10.1016/j.neuroimage.2019.06.017

Soares, J. M., Marques, P., Alves, V., & Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in Neuroscience*, *7*. doi: 10.3389/fnins.2013.00031

Song, J., Bergen, S. E., Kuja-Halkola, R., Larsson, H., Landén, M., & Lichtenstein, P. (2015). Bipolar disorder and its relation to major psychiatric disorders: A family-based study in the Swedish population. *Bipolar Disorders*, *17*(2), 184–93. doi: 10.1111/bdi.12242

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience*, *6*(3), 309–315. doi: 10.1038/nn1008

Spencer, T. J., Brown, A., Seidman, L. J., Valera, E. M., Makris, N., Lomedico, A., ... Biederman, J. (2013). Effect of Psychostimulants on Brain Structure and Function in ADHD: A Qualitative Literature Review of MRI-Based Neuroimaging Studies. *The Journal of clinical psychiatry*, *74*(9), 902–917. doi: 10.4088/JCP.12r08287

Takayanagi, Y., Takahashi, T., Orikabe, L., Masuda, N., Mozue, Y., Nakamura, K., ... Suzuki, M. (2010). Volume reduction and altered sulco-gyral pattern of the orbitofrontal cortex in first-episode schizophrenia. *Schizophrenia Research*, *121*(1), 55–65. doi: 10.1016/j.schres.2010.05.006

Tamnes, C. K., Herting, M. M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R. E., ... Mills, K. L. (2017). Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. *Journal of Neuroscience*, *37*(12), 3402–3412. doi: 10.1523/JNEUROSCI.3302-16.2017

Tamnes, C. K., Walhovd, K. B., Dale, A. M., Østby, Y., Grydeland, H., Richardson, G., ... Fjell, A. M. (2013). Brain development and aging: Overlapping and unique patterns of change. *NeuroImage*, *68*, 63–74. doi: 10.1016/j.neuroimage.2012.11.039

Team, R., & Others. (2015). RStudio: Integrated development for R. *RStudio, Inc. , Boston, MA URL http://www. rstudio. com*, *42*.

Teicher, M. H., Andersen, S. L., Polcari, A., Anderson, C. M., Navalta, C. P., & Kim, D. M. (2003). The neurobiological consequences of early stress and childhood maltreatment. *Neuroscience & Biobehavioral Reviews*, *27*(1), 33–44. doi: 10 .1016/S0149-7634(03)00007-1

Teicher, M. H., Dumont, N. L., Ito, Y., Vaituzis, C., Giedd, J. N., & Andersen, S. L. (2004). Childhood neglect is associated with reduced corpus callosum area. *Biological Psychiatry*, *56*(2), 80–85. doi: 10.1016/j.biopsych.2004.03.016

Thomason, M. E., Dennis, E. L., Joshi, A. A., Joshi, S. H., Dinov, I. D., Chang, C., ... Gotlib, I. H. (2011). Resting-state fMRI can reliably map neural networks in children. *NeuroImage*, *55*(1), 165–175. doi: 10.1016/j.neuroimage.2010.11.080

Thompson, C. A., Karelis, J., Middleton, F. A., Gentile, K., Coman, I. L., Radoeva, P. D., ... Kates, W. R. (2017). Associations between neurodevelopmental genes, neuroanatomy, and ultra high risk symptoms of psychosis in 22q11.2 deletion syndrome. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *174*(3), 295–314. doi: 10.1002/ajmg.b.32515

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., ... Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group (2014). The ENIGMA Consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, *8*(2), 153–182. doi: 10.1007/s11682-013-9269-5

Todd, R. D., Reich, W., Petti, T. A., Joshi, P., DePAULO, J. R., Nurnberger, J., & Reich, T. (1996). Psychiatric Diagnoses in the Child and Adolescent Members of Extended Families Identified through Adult Bipolar Affective Disorder Probands. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(5), 664–671. doi: 10.1097/00004583-199605000-00022

Touchette, E., Chollet, A., Galéra, C., Fombonne, E., Falissard, B., Boivin, M., & Melchior, M. (2012). Prior sleep problems predict internalising problems later in life. *Journal of Affective Disorders*, *143*(1-3), 166–171. doi: 10.1016/ j.jad.2012.05.049

Townsend, J. D., Bookheimer, S. Y., Foland-Ross, L. C., Moody, T. D., Eisenberger, N. I., Fischer, J. S., ... Altshuler, L. L. (2012). Deficits in inferior frontal cortex activation in euthymic bipolar disorder patients during a response inhibition task. *Bipolar Disorders*, *14*(4), 442. doi: 10.1111/j.1399-5618.2012.01020.x

Uher, R., Cumby, J., MacKenzie, L. E., Morash-Conway, J., Glover, J. M., Aylott, A., ... Alda, M. (2014). A familial risk enriched cohort as a platform for testing early interventions to prevent severe mental illness. *BMC Psychiatry*, *14*(1), 344. doi: 10.1186/s12888-014-0344-2

Uher, R., & Zwicker, A. (2017). Etiology in psychiatry: Embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry*, *16*(2), 121–129. doi: 10.1002/wps.20436

van Erp, T. G. M., Hibar, D. P., Rasmussen, J. M., Glahn, D. C., Pearlson, G. D., Andreassen, O. A., ... Turner, J. A. (2016). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Molecular Psychiatry*, *21*(4), 547–553. doi: 10.1038/mp.2015.63

van Erp, T. G. M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., ... Turner, J. A. (2018). Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biological Psychiatry*, *84*(9), 644–654. doi: 10.1016/j.biopsych.2018.04.023

Van Dijk, K. R., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, *59*(1), 431–438. doi: 10.1016/j.neuroimage.2011.07.044

Van Gestel, H., Franke, K., Petite, J., Slaney, C., Garnham, J., Helmick, C., ... Hajek, T. (2019). Brain age in bipolar disorders: Effects of lithium treatment. *Australian & New Zealand Journal of Psychiatry*, 000486741985781. doi: 10.1177/0004867419857814

Vetter, N. C., Steding, J., Jurk, S., Ripke, S., Mennigen, E., & Smolka, M. N. (2017). Reliability in adolescent fMRI within two years – a comparison of three tasks. *Scientific Reports*, *7*(1), 1–11. doi: 10.1038/s41598-017-02334-7

Vulser, H., Paillère Martinot, M.-L., Artiges, E., Miranda, R., Penttilä, J., Grimmer, Y., ... IMAGEN Consortium (2018). Early Variations in White Matter Microstructure and Depression Outcome in Adolescents With Subthreshold Depression. *The American Journal of Psychiatry*, *175*(12), 1255–1264. doi: 10.1176/appi.ajp.2018.17070825

Walhovd, K. B., Fjell, A. M., Giedd, J., Dale, A. M., & Brown, T. T. (2017). Through Thick and Thin: A Need to Reconcile Contradictory Results on Trajectories in Human Cortical Development. *Cerebral Cortex*, *27*(2), 1472–1481. doi: 10.1093/cercor/bhv301

Walton, E., Hibar, D. P., van Erp, T. G. M., Potkin, S. G., Roiz-Santiañez, R., Crespo-Facorro, B., ... Ehrlich, S. (2018). Prefrontal cortical thinning links to negative symptoms in schizophrenia via the ENIGMA consortium. *Psychological Medicine*, *48*(1), 82–94. doi: 10.1017/S0033291717001283

Wechsler, D. (2011). *WASI-II: Wechsler abbreviated scale of intelligence*. PsychCorp.

Weinberger, D. R. (1988). Schizophrenia and the frontal lobe. *Trends in Neurosciences*, *11*(8), 367–370. doi: 10.1016/0166-2236(88)90060-4

Weissman, M. M., Wickramaratne, P., Nomura, Y., Warner, V., Pilowsky, D., & Verdeli, H. (2006). Offspring of depressed parents: 20 years later. *The American Journal of Psychiatry*, *163*(6), 1001–1008. doi: 10.1176/ajp.2006.163.6.1001

Whitaker, K. J., Vértes, P. E., Romero-Garcia, R., Váša, F., Moutoussis, M., Prabhu, G., ... Bullmore, E. T. (2016). Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proceedings of the National Academy of Sciences*, *113*(32), 9105–9110. doi: 10.1073/pnas.1601745113

White, T., & Gottesman, I. (2012). Brain Connectivity and Gyrification as Endophenotypes for Schizophrenia: Weight of the Evidence. *Current Topics in Medicinal Chemistry*, *12*(21), 2393–2403. doi: 10.2174/156802612805289953

White, T., Su, S., Schmidt, M., Kao, C.-Y., & Sapiro, G. (2010). The development of gyrification in childhood and adolescence. *Brain and Cognition*, *72*(1), 36–45. doi: 10.1016/j.bandc.2009.10.009

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, *87*, 120–126. doi: 10.1016/j.neuroimage.2013.11.010

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., ... Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*(3), 1135–46. doi: 10.1016/j.neuroimage.2009.12.028

Winsberg, M. E., Sachs, N., Tate, D. L., Adalsteinsson, E., Spielman, D., & Ketter, T. A. (2000). Decreased dorsolateral prefrontal N-acetyl aspartate in bipolar disorder. *Biological Psychiatry*, *47*(6), 475–481.

Wise, T., Radua, J., Nortje, G., Cleare, A. J., Young, A. H., & Arnone, D. (2016). Voxel-Based Meta-Analytical Evidence of Structural Disconnectivity in Major Depression and Bipolar Disorder. *Biological Psychiatry*, *79*(4), 293–302. doi: 10.1016/j.biopsych.2015.03.004

Wise, T., Radua, J., Via, E., Cardoner, N., Abe, O., Adams, T. M., ... Arnone, D. (2017). Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: Evidence from voxel-based meta-analysis. *Molecular Psychiatry*, *22*(10), 1455–1463. doi: 10.1038/mp.2016.72

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods Ecol. Evol.*, *3*(1), 129–137.

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. doi: 10.1038/nn.4478

Wu, R. H., O'Donnell, T., Ulrich, M., Asghar, S. J., Hanstock, C. C., & Silverstone, P. H. (2004). Brain choline concentrations may not be altered in euthymic bipolar disorder patients chronically treated with either lithium or sodium valproate. *Annals of General Hospital Psychiatry*, *3*, 13. doi: 10.1186/1475-2832-3-13

Yucel, K., McKinnon, M. C., Taylor, V. H., Macdonald, K., Alda, M., Young, L. T., & MacQueen, G. M. (2007). Bilateral hippocampal volume increases after long-term lithium treatment in patients with bipolar disorder: A longitudinal MRI study. *Psychopharmacology*, *195*(3), 357–367. doi: 10.1007/s00213-007-0906-9

Zwicker, A., Denovan-Wright, E. M., & Uher, R. (2018). Gene–environment interplay in the etiology of psychosis. *Psychological Medicine*, *48*(12), 1925–1936. doi: 10.1017/S003329171700383X

Zwicker, A., Drobinin, V., MacKenzie, L. E., Howes Vallis, E., Patterson, V. C., Cumby, J., ... Uher, R. (2019). Affective lability in offspring of parents with major depressive disorder, bipolar disorder and schizophrenia. *European Child & Adolescent Psychiatry*. doi: 10.1007/s00787-019-01355-z

Zwicker, A., MacKenzie, L. E., Drobinin, V., Bagher, A. M., Howes Vallis, E., Propper, L., ... Uher, R. (2019). Neurodevelopmental and genetic determinants of exposure to adversity among youth at risk for mental illness. *Journal of Child Psychology and Psychiatry*, jcpp.13159. doi: 10.1111/jcpp.13159

Appendix A

Supplemental material

A.1 Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders

A subset of the participants has completed the Hamilton Depression Rating Scale (HAM-D) and Young Mania Rating Scale (YMRS) within one month of scanning. There were no statistically significant associations between the cortical measures and the clinical scales, with the greatest negative relationship being between right pars triangularis thickness and YMRS (Pearson r(28) = –.25, p = .19), see Supplement Figure A.1 below.

Figure A.1: Correlation heat map between right hemisphere pars triangularis structural measures and clinical scales of mania and depression. Based on 30 individuals with available data within 1 month of scanning.

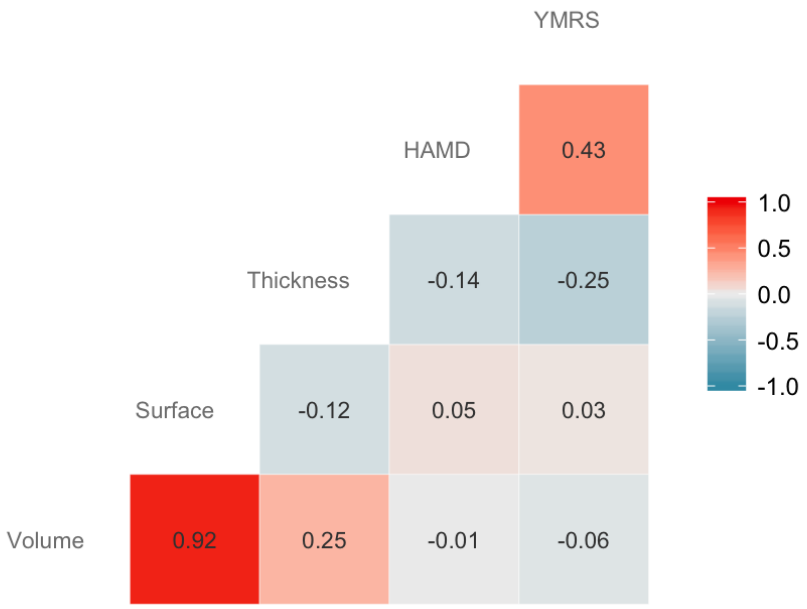
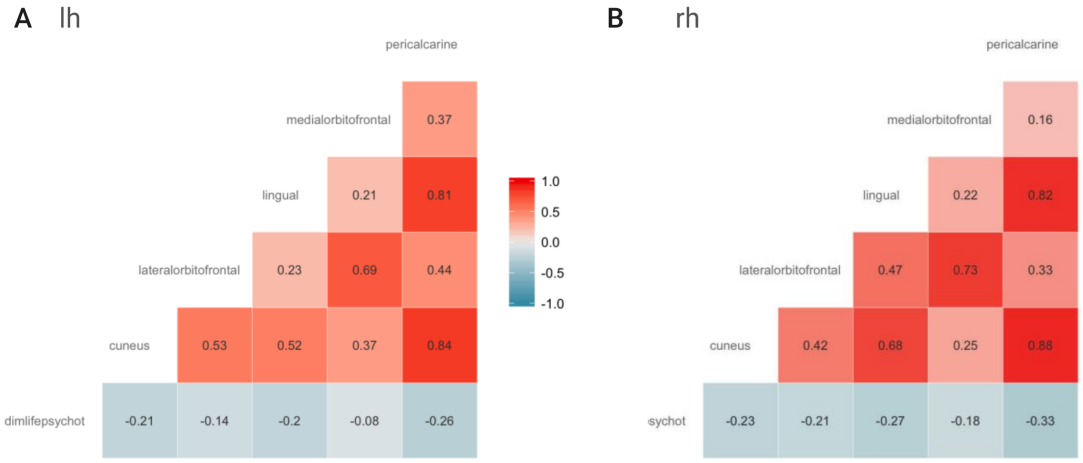


Figure A.2: Correlation matrix heat map between A. the left hemisphere and B. the right hemisphere ROIs and dimlifepsychot, a dimensional variable indexing psychotic symptoms. Based on 48 scans from individuals experiencing psychotic symptoms.



A.2 Psychotic symptoms are associated with lower cortical folding in youth at risk for mental illness

As a supplementary analysis, we examined psychotic symptoms from a dimensional perspective. We see a pattern of results similar to the categorical analyses, however there is a greater shift towards the prominence of occipital findings; pericalcarine β = -0.28, $p < 0.001$, cuneus β = -0.25, $p < 0.001$, lingual β = -0.24, $p = 0.001$. The lateral orbitofrontal (β = -0.15, $p = 0.04$) but not the medial orbitofrontal (β = -0.14, $p = 0.06$) gyrus remains significant after correcting for multiple testing. The effects are even larger when restricting analysis to the group with psychotic symptoms, suggesting that multiple definite psychotic symptoms are associated with even greater reduction in cortical folding (Supplement Figure A.2). Overall, dimensional capture of psychotic symptoms is negatively associated with the degree of cortical folding.

A.3 Reliability of multimodal MRI brain measures in youth at risk for mental illness

Table A.1: Grey matter volume reliability

		Left Hemisphere			Right Hemisphere		
ROI	**Overall ICC**	**2.5% CI**	**ICC**	**97.5% CI**	**2.5% CI**	**ICC**	**97.5% CI**
Banks of the Superior Temporal Sulcus	0.95	0.93	0.96	0.98	0.91	0.95	0.97
Caudal Anterior Cingulate	0.86	0.71	0.82	0.89	0.84	0.91	0.95
Caudal Middle Frontal	0.90	0.94	0.97	0.98	0.72	0.83	0.90
Cuneus	0.96	0.92	0.95	0.97	0.95	0.97	0.98
Entorhinal	0.79	0.72	0.83	0.90	0.60	0.75	0.85
Fusiform	0.96	0.90	0.94	0.97	0.95	0.97	0.98
Inferior Parietal	0.97	0.93	0.96	0.98	0.96	0.98	0.99
Inferior Temporal	0.98	0.97	0.98	0.99	0.95	0.97	0.98
Isthmus Cingulate	0.93	0.92	0.95	0.97	0.85	0.91	0.95
Lateral Occipital	0.96	0.91	0.95	0.97	0.94	0.97	0.98
Lateral Orbitofrontal	0.87	0.82	0.89	0.94	0.75	0.85	0.91
Lingual	0.97	0.96	0.97	0.99	0.92	0.96	0.97
Medial Orbitofrontal	0.88	0.70	0.82	0.89	0.89	0.94	0.96
Middle Temporal	0.96	0.95	0.97	0.98	0.93	0.96	0.98
Parahippocampal	0.90	0.85	0.91	0.95	0.80	0.88	0.93
Paracentral	0.93	0.92	0.95	0.97	0.84	0.91	0.94
Pars Opercularis	0.90	0.95	0.97	0.98	0.72	0.83	0.90
Pars Orbitalis	0.85	0.68	0.81	0.89	0.82	0.89	0.94
Pars Triangularis	0.94	0.92	0.95	0.97	0.86	0.92	0.95
Pericalcarine	0.95	0.91	0.94	0.97	0.91	0.95	0.97
Postcentral	0.96	0.97	0.98	0.99	0.88	0.93	0.96
Posterior Cingulate	0.93	0.88	0.93	0.96	0.88	0.93	0.96
Precentral	0.94	0.91	0.95	0.97	0.90	0.94	0.96
Precuneus	0.93	0.97	0.99	0.99	0.79	0.87	0.93
Rostral Anterior Cingulate	0.93	0.89	0.94	0.96	0.86	0.92	0.95
Rostral Middle Frontal	0.94	0.95	0.97	0.98	0.84	0.91	0.95
Superior Frontal	0.94	0.97	0.98	0.99	0.82	0.89	0.94
Superior Parietal	0.91	0.96	0.98	0.99	0.72	0.83	0.90
Superior Temporal	0.97	0.96	0.98	0.99	0.94	0.96	0.98
Supramarginal	0.92	0.98	0.99	0.99	0.76	0.86	0.92
Frontal Pole	0.68	0.64	0.78	0.87	0.36	0.58	0.74
Temporal Pole	0.51	0.23	0.47	0.66	0.33	0.55	0.72
Transverse Temporal	0.90	0.84	0.91	0.95	0.82	0.89	0.94
Insula	0.87	0.75	0.85	0.91	0.83	0.90	0.94

Table A.2: Cortical surface area reliability

		Left Hemisphere			Right Hemisphere		
ROI	**Overall ICC**	**2.5% CI**	**ICC**	**97.5% CI**	**2.5% CI**	**ICC**	**97.5% CI**
Banks of the Superior Temporal Sulcus	0.95	0.92	0.96	0.97	0.91	0.94	0.97
Caudal Anterior Cingulate	0.89	0.80	0.88	0.93	0.83	0.90	0.94
Caudal Middle Frontal	0.84	0.97	0.98	0.99	0.53	0.70	0.82
Cuneus	0.96	0.92	0.95	0.97	0.94	0.96	0.98
Entorhinal	0.74	0.55	0.71	0.83	0.62	0.77	0.86
Fusiform	0.95	0.87	0.93	0.96	0.95	0.97	0.98
Inferior Parietal	0.97	0.93	0.96	0.98	0.96	0.98	0.99
Inferior Temporal	0.95	0.91	0.95	0.97	0.92	0.96	0.97
Isthmus Cingulate	0.93	0.90	0.94	0.97	0.86	0.92	0.95
Lateral Occipital	0.97	0.92	0.96	0.97	0.97	0.98	0.99
Lateral Orbitofrontal	0.81	0.70	0.82	0.89	0.67	0.80	0.88
Lingual	0.96	0.94	0.97	0.98	0.91	0.95	0.97
Medial Orbitofrontal	0.77	0.45	0.64	0.78	0.82	0.89	0.94
Middle Temporal	0.97	0.95	0.97	0.98	0.94	0.97	0.98
Parahippocampal	0.88	0.84	0.91	0.95	0.77	0.86	0.92
Paracentral	0.95	0.93	0.96	0.98	0.91	0.94	0.97
Pars Opercularis	0.93	0.93	0.96	0.98	0.83	0.90	0.94
Pars Orbitalis	0.90	0.86	0.92	0.95	0.79	0.88	0.93
Pars Triangularis	0.92	0.92	0.95	0.97	0.81	0.89	0.93
Pericalcarine	0.94	0.87	0.92	0.95	0.93	0.96	0.98
Postcentral	0.91	0.98	0.99	0.99	0.73	0.84	0.90
Posterior Cingulate	0.90	0.82	0.89	0.94	0.85	0.91	0.95
Precentral	0.86	0.94	0.97	0.98	0.61	0.76	0.85
Precuneus	0.93	0.98	0.99	0.99	0.79	0.88	0.93
Rostral Anterior Cingulate	0.89	0.84	0.91	0.95	0.78	0.87	0.92
Rostral Middle Frontal	0.91	0.94	0.97	0.98	0.76	0.86	0.92
Superior Frontal	0.89	0.95	0.97	0.98	0.69	0.81	0.89
Superior Parietal	0.89	0.98	0.99	0.99	0.67	0.80	0.88
Superior Temporal	0.98	0.97	0.98	0.99	0.95	0.97	0.98
Supramarginal	0.92	0.98	0.99	0.99	0.74	0.84	0.91
Frontal Pole	0.69	0.51	0.69	0.81	0.53	0.70	0.82
Temporal Pole	0.56	0.45	0.65	0.78	0.23	0.47	0.66
Transverse Temporal	0.93	0.88	0.93	0.96	0.87	0.92	0.96
Insula	0.74	0.55	0.72	0.83	0.62	0.76	0.86

Table A.3: Cortical thickness reliability

		Left Hemisphere			Right Hemisphere		
ROI	**Overall ICC**	**2.5% CI**	**ICC**	**97.5% CI**	**2.5% CI**	**ICC**	**97.5% CI**
Banks of the Superior Temporal Sulcus	0.93	0.91	0.95	0.97	0.86	0.92	0.95
Caudal Anterior Cingulate	0.81	0.65	0.78	0.87	0.72	0.83	0.90
Caudal Middle Frontal	0.91	0.85	0.91	0.95	0.85	0.91	0.95
Cuneus	0.89	0.81	0.89	0.94	0.80	0.88	0.93
Entorhinal	0.61	0.43	0.63	0.77	0.39	0.60	0.75
Fusiform	0.87	0.83	0.90	0.94	0.73	0.84	0.90
Inferior Parietal	0.93	0.89	0.94	0.96	0.88	0.93	0.96
Inferior Temporal	0.87	0.76	0.85	0.91	0.80	0.88	0.93
Isthmus Cingulate	0.75	0.64	0.78	0.87	0.55	0.71	0.83
Lateral Occipital	0.87	0.79	0.87	0.93	0.76	0.86	0.92
Lateral Orbitofrontal	0.68	0.50	0.68	0.81	0.49	0.67	0.80
Lingual	0.84	0.73	0.83	0.90	0.76	0.85	0.91
Medial Orbitofrontal	0.62	0.34	0.56	0.73	0.48	0.67	0.80
Middle Temporal	0.85	0.81	0.89	0.93	0.68	0.81	0.89
Parahippocampal	0.88	0.79	0.88	0.93	0.79	0.88	0.93
Paracentral	0.86	0.79	0.88	0.93	0.76	0.85	0.91
Pars Opercularis	0.90	0.87	0.92	0.96	0.78	0.87	0.92
Pars Orbitalis	0.75	0.60	0.75	0.85	0.60	0.75	0.85
Pars Triangularis	0.92	0.87	0.92	0.95	0.87	0.92	0.95
Pericalcarine	0.73	0.54	0.71	0.83	0.60	0.75	0.85
Postcentral	0.80	0.91	0.95	0.97	0.47	0.66	0.79
Posterior Cingulate	0.87	0.76	0.85	0.91	0.80	0.88	0.93
Precentral	0.82	0.85	0.91	0.95	0.56	0.73	0.83
Precuneus	0.93	0.88	0.93	0.96	0.87	0.92	0.96
Rostral Anterior Cingulate	0.74	0.60	0.75	0.85	0.56	0.72	0.83
Rostral Middle Frontal	0.93	0.85	0.91	0.95	0.91	0.95	0.97
Superior Frontal	0.95	0.91	0.95	0.97	0.90	0.94	0.97
Superior Parietal	0.92	0.89	0.94	0.96	0.84	0.90	0.94
Superior Temporal	0.92	0.87	0.92	0.96	0.85	0.91	0.95
Supramarginal	0.93	0.90	0.94	0.97	0.85	0.91	0.95
Frontal Pole	0.69	0.62	0.76	0.86	0.42	0.62	0.77
Temporal Pole	0.40	0.12	0.38	0.60	0.16	0.41	0.62
Transverse Temporal	0.82	0.71	0.82	0.90	0.71	0.83	0.90
Insula	0.63	0.46	0.66	0.79	0.41	0.61	0.76

Table A.4: Local gyrification index reliability

		Left Hemisphere			Right Hemisphere		
ROI	**Overall ICC**	**2.5% CI**	**ICC**	**97.5% CI**	**2.5% CI**	**ICC**	**97.5% CI**
Banks of the Superior Temporal Sulcus	0.85	0.68	0.81	0.89	0.81	0.89	0.94
Caudal Anterior Cingulate	0.82	0.68	0.81	0.89	0.73	0.84	0.91
Caudal Middle Frontal	0.93	0.91	0.95	0.97	0.87	0.92	0.96
Cuneus	0.90	0.87	0.92	0.95	0.79	0.87	0.93
Entorhinal	0.68	0.45	0.65	0.78	0.54	0.71	0.82
Fusiform	0.83	0.65	0.79	0.87	0.80	0.88	0.93
Inferior Parietal	0.92	0.82	0.90	0.94	0.91	0.95	0.97
Inferior Temporal	0.78	0.53	0.70	0.82	0.76	0.86	0.92
Isthmus Cingulate	0.83	0.70	0.82	0.89	0.76	0.85	0.91
Lateral Occipital	0.92	0.84	0.90	0.94	0.89	0.94	0.96
Lateral Orbitofrontal	0.79	0.65	0.78	0.87	0.66	0.79	0.88
Lingual	0.89	0.81	0.89	0.93	0.83	0.90	0.94
Medial Orbitofrontal	0.73	0.54	0.71	0.82	0.59	0.75	0.85
Middle Temporal	0.81	0.60	0.75	0.85	0.80	0.88	0.93
Parahippocampal	0.79	0.64	0.78	0.87	0.66	0.79	0.88
Paracentral	0.84	0.80	0.88	0.93	0.68	0.81	0.88
Pars Opercularis	0.92	0.87	0.92	0.95	0.85	0.91	0.95
Pars Orbitalis	0.78	0.61	0.76	0.85	0.67	0.80	0.88
Pars Triangularis	0.88	0.79	0.87	0.93	0.82	0.89	0.94
Pericalcarine	0.91	0.89	0.94	0.96	0.81	0.89	0.94
Postcentral	0.93	0.86	0.92	0.95	0.89	0.93	0.96
Posterior Cingulate	0.82	0.71	0.82	0.89	0.71	0.82	0.89
Precentral	0.95	0.90	0.94	0.97	0.92	0.95	0.97
Precuneus	0.89	0.84	0.90	0.94	0.79	0.87	0.93
Rostral Anterior Cingulate	0.80	0.68	0.81	0.89	0.67	0.80	0.88
Rostral Middle Frontal	0.90	0.84	0.90	0.94	0.83	0.90	0.94
Superior Frontal	0.88	0.82	0.89	0.94	0.78	0.87	0.92
Superior Parietal	0.89	0.86	0.92	0.95	0.77	0.86	0.92
Superior Temporal	0.93	0.89	0.94	0.96	0.86	0.92	0.95
Supramarginal	0.91	0.82	0.89	0.94	0.87	0.93	0.96
Frontal Pole	0.64	0.41	0.62	0.76	0.48	0.66	0.79
Temporal Pole	0.66	0.43	0.63	0.77	0.50	0.68	0.81
Transverse Temporal	0.92	0.87	0.92	0.96	0.85	0.91	0.95
Insula	0.85	0.77	0.86	0.92	0.73	0.84	0.91

Table A.5: White matter volume reliability

ROI	Hemisphere	2.5% CI	ICC	97.5% CI	Classification
Anterior Thalamic Radiation	lh	0.96	0.98	0.99	Excellent
Anterior Thalamic Radiation	rh	0.97	0.98	0.99	Excellent
Cingulum Cingulate Gyrus	lh	0.99	0.99	1.00	Excellent
Cingulum Cingulate Gyrus	rh	0.98	0.99	0.99	Excellent
Cingulum Hippocampus	lh	0.93	0.96	0.98	Excellent
Cingulum Hippocampus	rh	0.95	0.97	0.98	Excellent
Corticospinal Tract	lh	0.96	0.98	0.99	Excellent
Corticospinal Tract	rh	0.97	0.98	0.99	Excellent
Forceps Major	both	0.97	0.98	0.99	Excellent
Forceps Minor	both	0.95	0.97	0.98	Excellent
Inferior Fronto-Occipital Fasciculus	lh	0.95	0.97	0.98	Excellent
Inferior Fronto-Occipital Fasciculus	rh	0.96	0.98	0.99	Excellent
Inferior Longitudinal Fasciculus	lh	0.95	0.97	0.98	Excellent
Inferior Longitudinal Fasciculus	rh	0.97	0.98	0.99	Excellent
Superior Longitudinal Fasciculus Temporal	lh	0.98	0.99	0.99	Excellent
Superior Longitudinal Fasciculus Temporal	rh	0.97	0.98	0.99	Excellent
Superior Longitudinal Fasciculus	lh	0.98	0.99	0.99	Excellent
Superior Longitudinal Fasciculus	rh	0.99	0.99	1.00	Excellent
Uncinate Fasciculus	lh	0.97	0.98	0.99	Excellent
Uncinate Fasciculus	rh	0.96	0.97	0.99	Excellent

Table A.6: Fractional anisotropy (FA) reliability

ROI	**Hemisphere**	**2.5% CI**	**ICC**	**97.5% CI**	**Classification**
Anterior Thalamic Radiation	lh	0.73	0.84	0.90	Excellent
Anterior Thalamic Radiation	rh	0.73	0.84	0.90	Excellent
Cingulum Cingulate Gyrus	lh	0.81	0.88	0.93	Excellent
Cingulum Cingulate Gyrus	rh	0.69	0.81	0.89	Excellent
Cingulum Hippocampus	lh	0.88	0.93	0.96	Excellent
Cingulum Hippocampus	rh	0.72	0.83	0.90	Excellent
Corticospinal Tract	lh	0.74	0.85	0.91	Excellent
Corticospinal Tract	rh	0.69	0.81	0.89	Excellent
Forceps Major	both	0.90	0.94	0.97	Excellent
Forceps Minor	both	0.61	0.76	0.86	Excellent
Inferior Fronto-Occipital Fasciculus	lh	0.84	0.91	0.95	Excellent
Inferior Fronto-Occipital Fasciculus	rh	0.83	0.90	0.94	Excellent
Inferior Longitudinal Fasciculus	lh	0.91	0.95	0.97	Excellent
Inferior Longitudinal Fasciculus	rh	0.88	0.93	0.96	Excellent
Superior Longitudinal Fasciculus Temporal	lh	0.89	0.94	0.96	Excellent
Superior Longitudinal Fasciculus Temporal	rh	0.90	0.94	0.97	Excellent
Superior Longitudinal Fasciculus	lh	0.91	0.95	0.97	Excellent
Superior Longitudinal Fasciculus	rh	0.89	0.94	0.96	Excellent
Uncinate Fasciculus	lh	0.65	0.78	0.87	Excellent
Uncinate Fasciculus	rh	0.69	0.81	0.89	Excellent

Table A.7: Radial diffusivity (RD) reliability

ROI	**Hemisphere**	**2.5% CI**	**ICC**	**97.5% CI**	**Classification**
Anterior Thalamic Radiation	lh	0.77	0.86	0.92	Excellent
Anterior Thalamic Radiation	rh	0.80	0.88	0.93	Excellent
Cingulum Cingulate Gyrus	lh	0.77	0.86	0.92	Excellent
Cingulum Cingulate Gyrus	rh	0.58	0.73	0.84	Good
Cingulum Hippocampus	lh	0.57	0.73	0.84	Good
Cingulum Hippocampus	rh	0.52	0.69	0.81	Good
Corticospinal Tract	lh	0.59	0.74	0.84	Excellent
Corticospinal Tract	rh	0.53	0.70	0.82	Good
Forceps Major	both	0.90	0.94	0.97	Excellent
Forceps Minor	both	0.87	0.92	0.96	Excellent
Inferior Fronto-Occipital Fasciculus	lh	0.79	0.87	0.93	Excellent
Inferior Fronto-Occipital Fasciculus	rh	0.80	0.88	0.93	Excellent
Inferior Longitudinal Fasciculus	lh	0.82	0.89	0.94	Excellent
Inferior Longitudinal Fasciculus	rh	0.67	0.80	0.88	Excellent
Superior Longitudinal Fasciculus Temporal	lh	0.85	0.91	0.95	Excellent
Superior Longitudinal Fasciculus Temporal	rh	0.85	0.91	0.95	Excellent
Superior Longitudinal Fasciculus	lh	0.87	0.92	0.96	Excellent
Superior Longitudinal Fasciculus	rh	0.83	0.90	0.94	Excellent
Uncinate Fasciculus	lh	0.70	0.82	0.89	Excellent
Uncinate Fasciculus	rh	0.63	0.77	0.86	Excellent

Table A.8: Mean diffusivity (MD) reliability

ROI	Hemisphere	2.5% CI	ICC	97.5% CI	Classification
Anterior Thalamic Radiation	lh	0.79	0.87	0.93	Excellent
Anterior Thalamic Radiation	rh	0.81	0.89	0.93	Excellent
Cingulum Cingulate Gyrus	lh	0.69	0.81	0.89	Excellent
Cingulum Cingulate Gyrus	rh	0.53	0.70	0.82	Good
Cingulum Hippocampus	lh	0.46	0.65	0.79	Good
Cingulum Hippocampus	rh	0.43	0.63	0.77	Good
Corticospinal Tract	lh	0.52	0.69	0.81	Good
Corticospinal Tract	rh	0.48	0.66	0.79	Good
Forceps Major	both	0.89	0.93	0.96	Excellent
Forceps Minor	both	0.85	0.91	0.95	Excellent
Inferior Fronto-Occipital Fasciculus	lh	0.74	0.84	0.91	Excellent
Inferior Fronto-Occipital Fasciculus	rh	0.78	0.87	0.92	Excellent
Inferior Longitudinal Fasciculus	lh	0.75	0.85	0.91	Excellent
Inferior Longitudinal Fasciculus	rh	0.59	0.74	0.84	Excellent
Superior Longitudinal Fasciculus Temporal	lh	0.82	0.89	0.94	Excellent
Superior Longitudinal Fasciculus Temporal	rh	0.81	0.88	0.93	Excellent
Superior Longitudinal Fasciculus	lh	0.84	0.91	0.95	Excellent
Superior Longitudinal Fasciculus	rh	0.80	0.88	0.93	Excellent
Uncinate Fasciculus	lh	0.69	0.81	0.89	Excellent
Uncinate Fasciculus	rh	0.60	0.75	0.85	Excellent

Table A.9: Axial diffusivity (AD) reliability

ROI	**Hemisphere**	**2.5% CI**	**ICC**	**97.5% CI**	**Classification**
Anterior Thalamic Radiation	lh	0.84	0.90	0.94	Excellent
Anterior Thalamic Radiation	rh	0.83	0.90	0.94	Excellent
Cingulum Cingulate Gyrus	lh	0.50	0.68	0.81	Good
Cingulum Cingulate Gyrus	rh	0.63	0.77	0.86	Excellent
Cingulum Hippocampus	lh	0.48	0.67	0.80	Good
Cingulum Hippocampus	rh	0.30	0.53	0.70	Fair
Corticospinal Tract	lh	0.43	0.63	0.77	Good
Corticospinal Tract	rh	0.47	0.66	0.79	Good
Forceps Major	both	0.88	0.93	0.96	Excellent
Forceps Minor	both	0.76	0.86	0.92	Excellent
Inferior Fronto-Occipital Fasciculus	lh	0.70	0.82	0.89	Excellent
Inferior Fronto-Occipital Fasciculus	rh	0.80	0.88	0.93	Excellent
Inferior Longitudinal Fasciculus	lh	0.63	0.77	0.86	Excellent
Inferior Longitudinal Fasciculus	rh	0.55	0.72	0.83	Good
Superior Longitudinal Fasciculus Temporal	lh	0.70	0.82	0.89	Excellent
Superior Longitudinal Fasciculus Temporal	rh	0.70	0.82	0.89	Excellent
Superior Longitudinal Fasciculus	lh	0.77	0.86	0.92	Excellent
Superior Longitudinal Fasciculus	rh	0.72	0.83	0.90	Excellent
Uncinate Fasciculus	lh	0.67	0.80	0.88	Excellent
Uncinate Fasciculus	rh	0.60	0.75	0.85	Excellent

Table A.10: Scan-rescan ICC averaged across the Desikan atlas regions for common structural measures. A/B scans (N=100) from the main results. C/D ICC from a subset of scans (N=46) collected an average of 14 months after the A/B scans. ICC = intraclass correlation coefficient. LGI = local gyrification index. CI = confidence interval.

Measure	**A/B ICC**	**C/D ICC**
Cortical Grey Matter Volume	0.90, 95% CI [0.84, 0.94]	0.92, 95% CI [0.83, 0.96]
Cortical Surface Area	0.89, 95% CI [0.82, 0.93].	0.92, 95% CI [0.83, 0.96]
Cortical Thickness	0.82, 95% CI [0.71, 0.89].	0.80, 95% CI [0.61, 0.91]
Cortical Folding (LGI)	0.85, 95% CI [0.75, 0.91]	0.83, 95% CI [0.66, 0.92]

In addition to the 100 scans from 50 individuals reported on in the manuscript (timepoints A/B), we have collected an additional 46 scans from 23 individuals (timepoints C/D). These scans were collected an average of 14 months after the A/B scans and allow us to address the generalizability of the findings to the same scanner over a year later.

Overall, conducting the analysis in an identical way to the original sample, we again find excellent reliability in the subset of the same individuals scanned 14 months later. Furthermore, supplement table 10 shows the high consistency of the results.

A.4 Brain Age Supplement

Table A.11: Scans across time points. The study design includes short term reliability (weeks) and stability scans as well as longitudinal follow up (months/years).

Scan time	**N**	**Description**
A	150	Baseline scan
B	49	Short term repeat of A
C	86	Longitudinal follow up 1
D	21	Short term repeat of C
E	24	Longitudinal follow up 2
F	3	Short term repeat of E
G	4	Longitudinal follow up 3
H	1	Short term repeat of G

Table A.12: List of features used in model training and prediction

FS_InterCranial_Vol
FS_BrainSeg_Vol
FS_BrainSeg_Vol_No_Vent
FS_BrainSeg_Vol_No_Vent_Surf
FS_LCort_GM_Vol
FS_RCort_GM_Vol
FS_TotCort_GM_Vol
FS_SubCort_GM_Vol
FS_Total_GM_Vol
FS_SupraTentorial_Vol
FS_SupraTentorial_Vol_No_Vent
FS_SupraTentorial_No_Vent_Voxel_Count
FS_Mask_Vol
FS_BrainSegVol_eTIV_Ratio
FS_MaskVol_eTIV_Ratio
FS_L_LatVent_Vol
FS_L_InfLatVent_Vol
FS_L_Cerebellum_WM_Vol
FS_L_Cerebellum_Cort_Vol
FS_L_ThalamusProper_Vol
FS_L_Caudate_Vol
FS_L_Putamen_Vol
FS_L_Pallidum_Vol
FS_3rdVent_Vol
FS_4thVent_Vol
FS_BrainStem_Vol
FS_L_Hippo_Vol
FS_L_Amygdala_Vol
FS_CSF_Vol
FS_L_AccumbensArea_Vol
FS_L_VentDC_Vol
FS_L_Vessel_Vol
FS_L_ChoroidPlexus_Vol
FS_R_LatVent_Vol
FS_R_InfLatVent_Vol
FS_R_Cerebellum_WM_Vol
FS_R_ThalamusProper_Vol
FS_R_Caudate_Vol
FS_R_Putamen_Vol
FS_R_Pallidum_Vol
FS_R_Hippo_Vol
FS_R_Amygdala_Vol
FS_R_AccumbensArea_Vol
FS_R_VentDC_Vol
FS_R_Vessel_Vol
FS_R_ChoroidPlexus_Vol
FS_OpticChiasm_Vol
FS_CC_Posterior_Vol
FS_CC_MidPosterior_Vol
FS_CC_Central_Vol
FS_CC_MidAnterior_Vol
FS_CC_Anterior_Vol
FS_L_Bankssts_Area
FS_L_Caudalanteriorcingulate_Area
FS_L_Caudalmiddlefrontal_Area
FS_L_Cuneus_Area
FS_L_Entorhinal_Area
FS_L_Fusiform_Area
FS_L_Inferiorparietal_Area
FS_L_Inferiortemporal_Area
FS_L_Isthmuscingulate_Area
FS_L_Lateraloccipital_Area
FS_L_Lateralorbitofrontal_Area
FS_L_Lingual_Area
FS_L_Medialorbitofrontal_Area
FS_L_Middletemporal_Area
FS_L_Parahippocampal_Area
FS_L_Paracentral_Area
FS_L_Parsopercularis_Area
FS_L_Parsorbitalis_Area
FS_L_Parstriangularis_Area
FS_L_Pericalcarine_Area
FS_L_Postcentral_Area
FS_L_Posteriorcingulate_Area
FS_L_Precentral_Area
FS_L_Precuneus_Area
FS_L_Rostralanteriorcingulate_Area
FS_L_Rostralmiddlefrontal_Area
FS_L_Superiorfrontal_Area
FS_L_Superiorparietal_Area
FS_L_Superiortemporal_Area
FS_L_Supramarginal_Area
FS_L_Frontalpole_Area
FS_L_Temporalpole_Area
FS_L_Transversetemporal_Area
FS_L_Insula_Area
FS_R_Bankssts_Area
FS_R_Caudalanteriorcingulate_Area
FS_R_Caudalmiddlefrontal_Area
FS_R_Cuneus_Area
FS_R_Entorhinal_Area
FS_R_Fusiform_Area
FS_R_Inferiorparietal_Area
FS_R_Inferiortemporal_Area
FS_R_Isthmuscingulate_Area
FS_R_Lateraloccipital_Area
FS_R_Lateralorbitofrontal_Area
FS_R_Lingual_Area
FS_R_Medialorbitofrontal_Area
FS_R_Middletemporal_Area
FS_R_Parahippocampal_Area
FS_R_Paracentral_Area
FS_R_Parsopercularis_Area
FS_R_Parsorbitalis_Area
FS_R_Parstriangularis_Area
FS_R_Pericalcarine_Area
FS_R_Postcentral_Area
FS_R_Posteriorcingulate_Area
FS_R_Precentral_Area
FS_R_Precuneus_Area
FS_R_Rostralanteriorcingulate_Area
FS_R_Rostralmiddlefrontal_Area
FS_R_Superiorfrontal_Area
FS_R_Superiorparietal_Area
FS_R_Superiortemporal_Area
FS_R_Supramarginal_Area
FS_R_Frontalpole_Area
FS_R_Temporalpole_Area
FS_R_Transversetemporal_Area
FS_R_Insula_Area
FS_L_Bankssts_GrayVol
FS_L_Caudalanteriorcingulate_GrayVol
FS_L_Caudalmiddlefrontal_GrayVol
FS_L_Cuneus_GrayVol
FS_L_Entorhinal_GrayVol
FS_L_Fusiform_GrayVol
FS_L_Inferiorparietal_GrayVol
FS_L_Inferiortemporal_GrayVol
FS_L_Isthmuscingulate_GrayVol
FS_L_Lateraloccipital_GrayVol
FS_L_Lateralorbitofrontal_GrayVol
FS_L_Lingual_GrayVol
FS_L_Medialorbitofrontal_GrayVol
FS_L_Middletemporal_GrayVol
FS_L_Parahippocampal_GrayVol
FS_L_Paracentral_GrayVol
FS_L_Parsopercularis_GrayVol
FS_L_Parsorbitalis_GrayVol
FS_L_Parstriangularis_GrayVol
FS_L_Pericalcarine_GrayVol
FS_L_Postcentral_GrayVol
FS_L_Posteriorcingulate_GrayVol
FS_L_Precentral_GrayVol
FS_L_Precuneus_GrayVol
FS_L_Rostralanteriorcingulate_GrayVol
FS_L_Rostralmiddlefrontal_GrayVol
FS_L_Superiorfrontal_GrayVol
FS_L_Superiorparietal_GrayVol
FS_L_Superiortemporal_GrayVol
FS_L_Supramarginal_GrayVol
FS_L_Frontalpole_GrayVol
FS_L_Temporalpole_GrayVol
FS_L_Transversetemporal_GrayVol
FS_L_Insula_GrayVol
FS_R_Bankssts_GrayVol
FS_R_Caudalanteriorcingulate_GrayVol
FS_R_Caudalmiddlefrontal_GrayVol
FS_R_Cuneus_GrayVol
FS_R_Entorhinal_GrayVol
FS_R_Fusiform_GrayVol
FS_R_Inferiorparietal_GrayVol
FS_R_Inferiortemporal_GrayVol
FS_R_Isthmuscingulate_GrayVol
FS_R_Lateraloccipital_GrayVol
FS_R_Lateralorbitofrontal_GrayVol
FS_R_Lingual_GrayVol
FS_R_Medialorbitofrontal_GrayVol
FS_R_Middletemporal_GrayVol
FS_R_Parahippocampal_GrayVol
FS_R_Paracentral_GrayVol
FS_R_Parsopercularis_GrayVol
FS_R_Parsorbitalis_GrayVol
FS_R_Parstriangularis_GrayVol
FS_R_Pericalcarine_GrayVol
FS_R_Postcentral_GrayVol
FS_R_Posteriorcingulate_GrayVol
FS_R_Precentral_GrayVol
FS_R_Precuneus_GrayVol
FS_R_Rostralanteriorcingulate_GrayVol
FS_R_Rostralmiddlefrontal_GrayVol
FS_R_Superiorfrontal_GrayVol
FS_R_Superiorparietal_GrayVol
FS_R_Superiortemporal_GrayVol
FS_R_Supramarginal_GrayVol
FS_R_Frontalpole_GrayVol
FS_R_Temporalpole_GrayVol
FS_R_Transversetemporal_GrayVol
FS_R_Insula_GrayVol

Figure A.3: Neuroanatomical contribution to age prediction visualised with a variable importance plot.

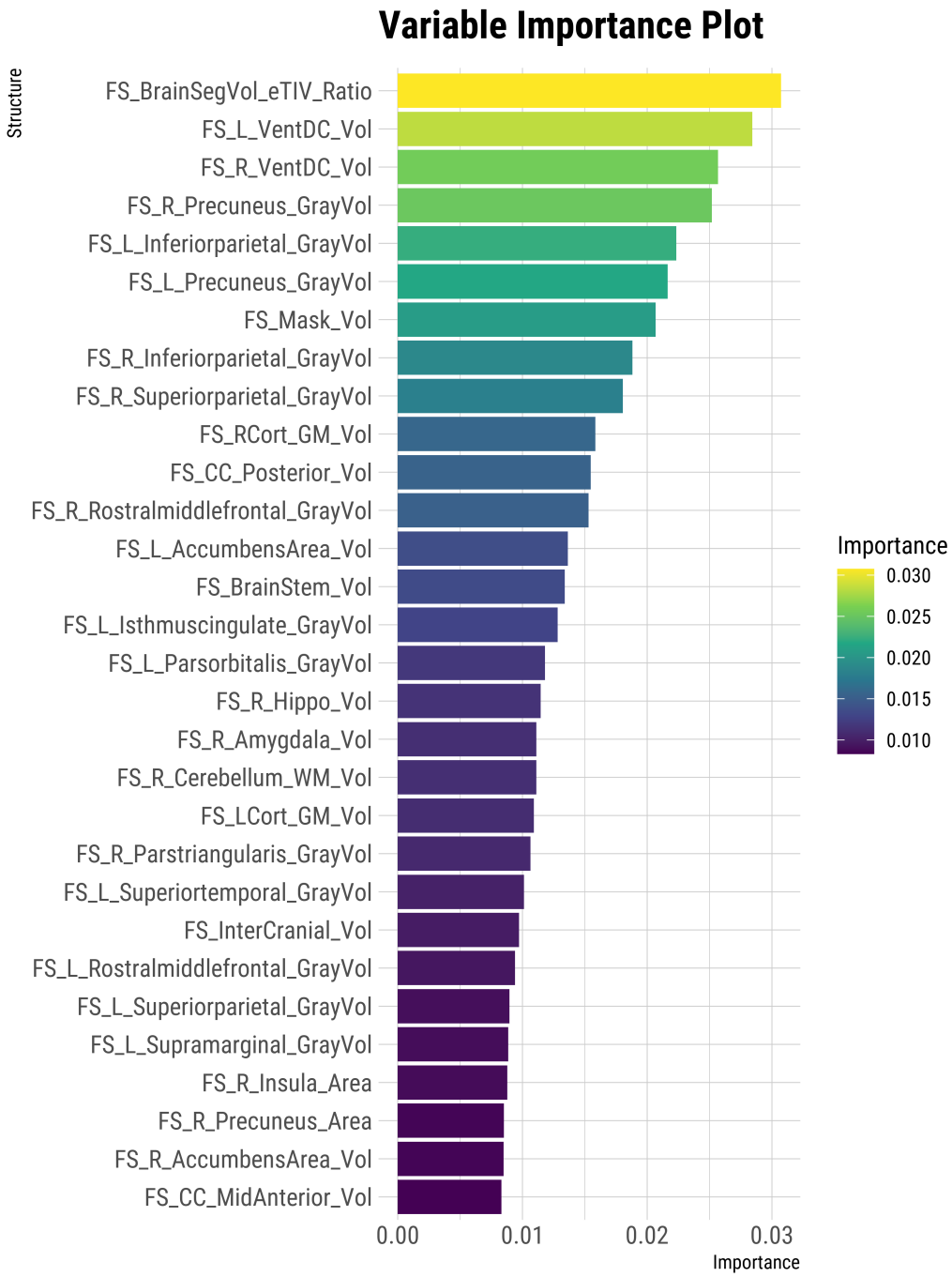


Figure A.4: Results from 3 primary brain age dimensional models and their sensitivity analyses.

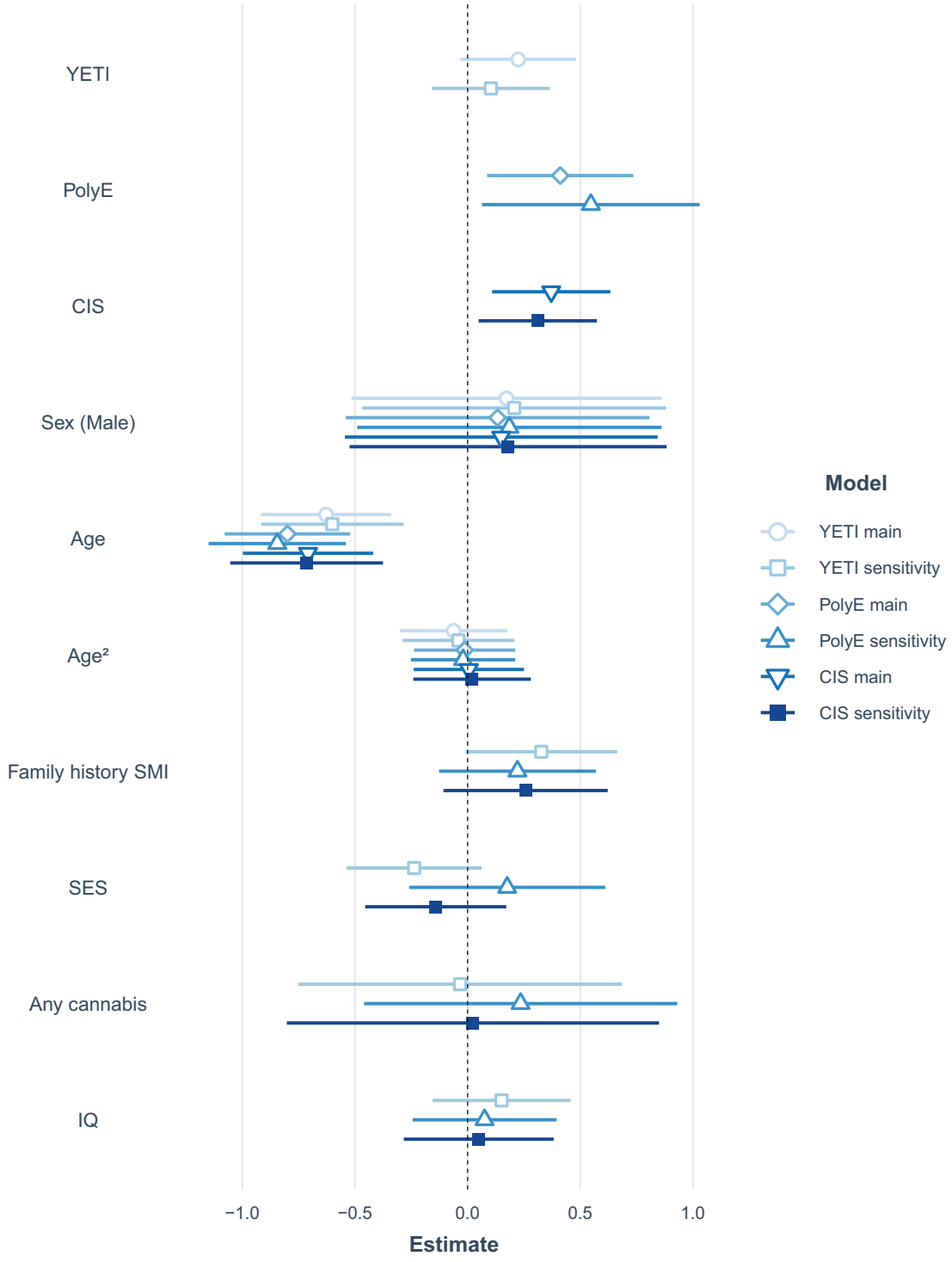


Table A.13: Parsnip standardized parameter names with their matching names in the underlying xgboost engine. Engine default values in brackets. Table and additional documentation available on the *tidymodels website*.

parsnip	**xgboost**
tree_depth	max_depth (6)
trees	nrounds (15)
learn_rate	eta (0.3)
mtry	colsample_bytree (1)
min_n	min_child_weight (1)
loss_reduction	gamma (0)
sample_size	subsample (1)
stop_iter	early_stop

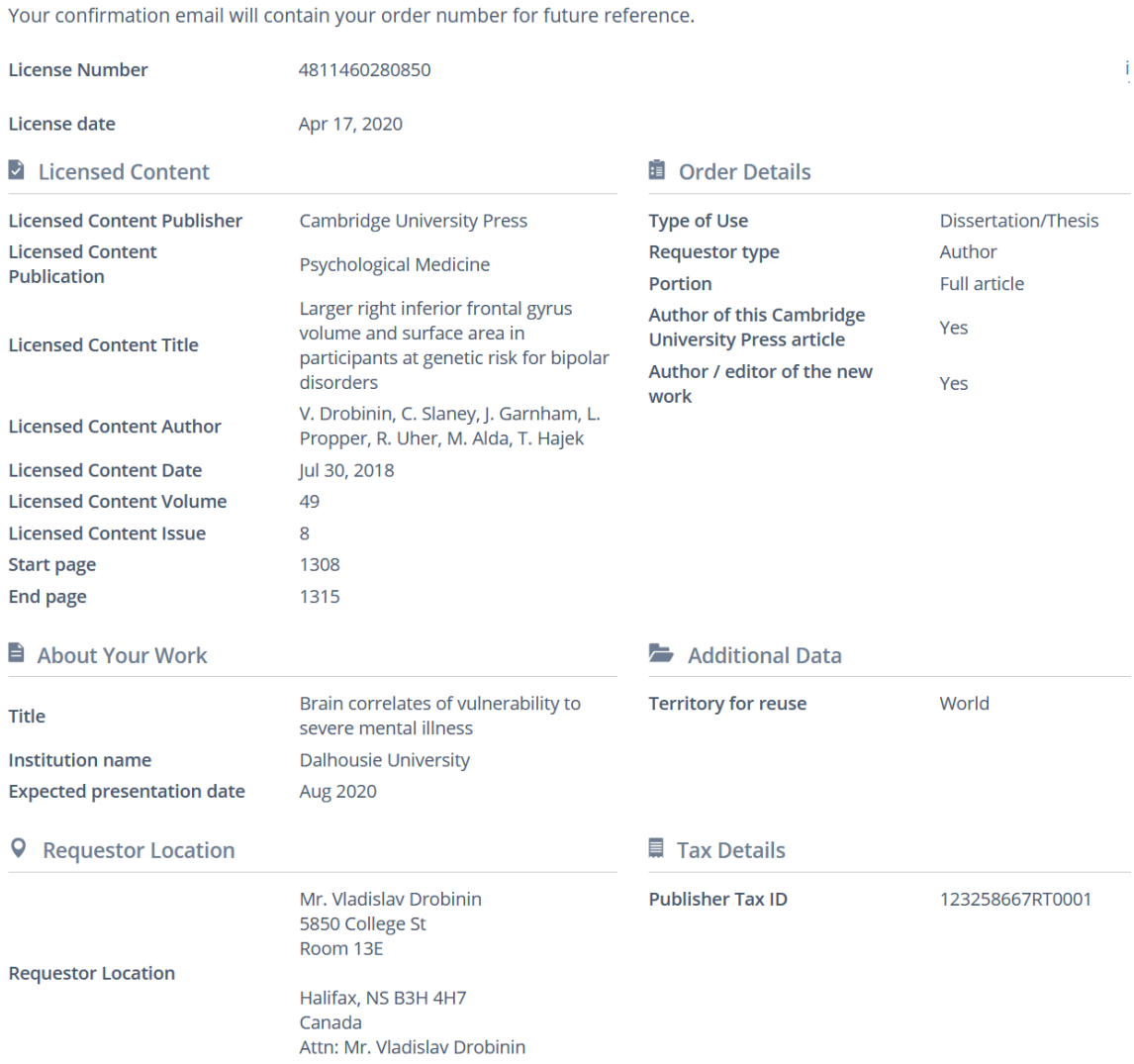
Table A.14: Parameters of top 10 best (lowest MAE) XGBoost models in cross-validation. The final model was selected by lowest tree depth within a standard error of best absolute performing model. Final model parameters in bold below.

mtry	**min_n**	**tree_depth**	**learn_rate**	**loss_reduction**	**sample_size**	**.metric**	**mean (MAE)**	**n**	**std_err**
36	17	9	0.015337	1.610e-07	0.3169	mae	1.523	100	0.008356
62	36	10	0.016105	2.353e-05	0.5631	mae	1.523	100	0.008136
88	22	13	0.014384	5.690e-06	0.5778	mae	1.525	100	0.008075
61	**12**	**6**	**0.013062**	**8.068e-04**	**0.4874**	**mae**	**1.526**	**100**	**0.007807**
176	5	10	0.011077	9.034e+00	0.2078	mae	1.528	100	0.008351
177	10	10	0.006773	3.157e-06	0.4177	mae	1.530	100	0.007813
92	14	11	0.005258	3.304e-06	0.4001	mae	1.535	100	0.008140
44	16	4	0.019228	3.387e-09	0.7696	mae	1.535	100	0.008000
186	17	11	0.026158	1.057e-06	0.3057	mae	1.535	100	0.008781

Appendix B

Copyright permissions

Figure B.1: License to include "Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders"



Your confirmation email will contain your order number for future reference.

License Number	4811460280850
License date	Apr 17, 2020

Licensed Content

Licensed Content Publisher	Cambridge University Press
Licensed Content Publication	Psychological Medicine
Licensed Content Title	Larger right inferior frontal gyrus volume and surface area in participants at genetic risk for bipolar disorders
Licensed Content Author	V. Drobinin, C. Slaney, J. Garnham, L. Propper, R. Uher, M. Alda, T. Hajek
Licensed Content Date	Jul 30, 2018
Licensed Content Volume	49
Licensed Content Issue	8
Start page	1308
End page	1315

Order Details

Type of Use	Dissertation/Thesis
Requestor type	Author
Portion	Full article
Author of this Cambridge University Press article	Yes
Author / editor of the new work	Yes

About Your Work

Title	Brain correlates of vulnerability to severe mental illness
Institution name	Dalhousie University
Expected presentation date	Aug 2020

Additional Data

Territory for reuse	World

Requestor Location

Requestor Location	Mr. Vladislav Drobinin 5850 College St Room 13E Halifax, NS B3H 4H7 Canada Attn: Mr. Vladislav Drobinin

Tax Details

Publisher Tax ID	123258667RT0001

Figure B.2: License to include "Psychotic symptoms are associated with lower cortical folding in youth at risk for mental illness license"

From: Access Copyright Permissions Group
Sent: April 23, 2020 10:36 AM
To: Vlad Drobinin
Subject: Re: Inclusion of authored article in PhD Thesis

Hi Vlad,

Thanks for your email. The Canadian Medical Association does not charge for requests from authors wishing to re-use their own material. As you are an original author and wish to re-use the requested material, the normal fees for use have been waived and you may proceed to use the material as noted. Please use the following copyright statement:

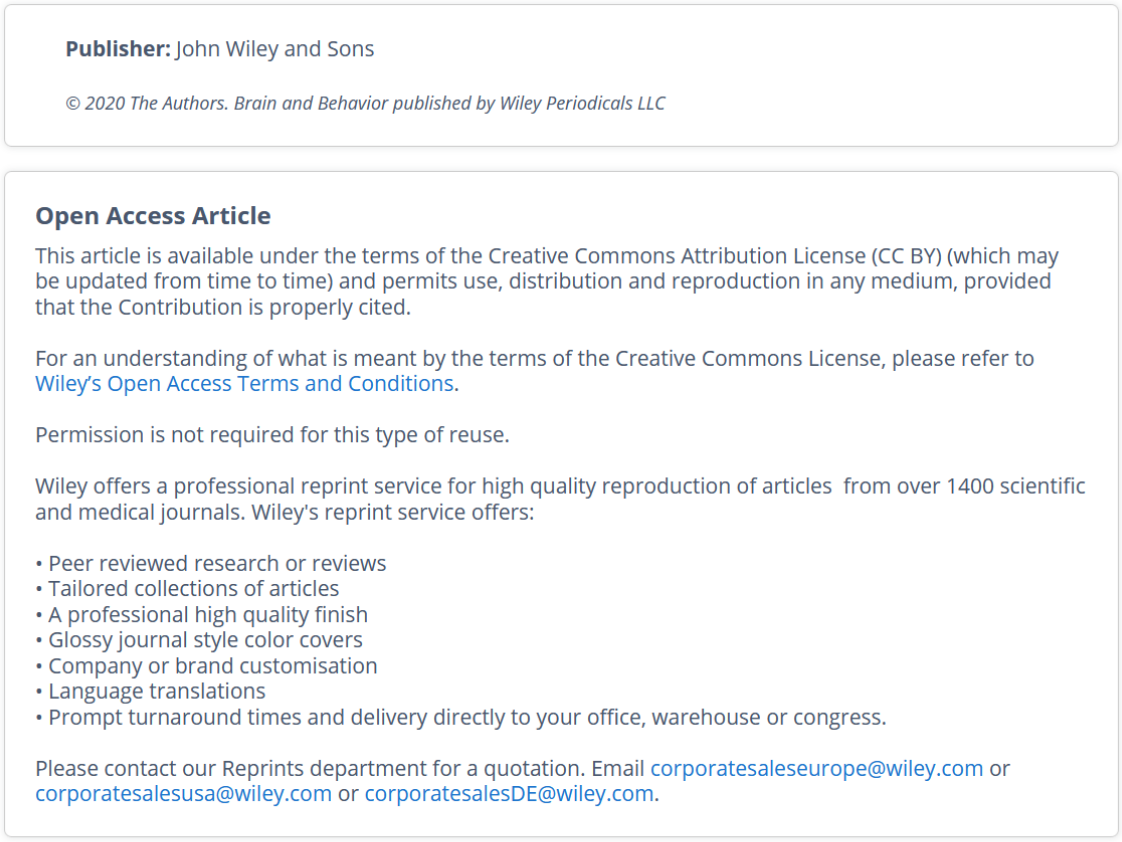
Reprinted from (insert author names) (insert article and figure title) Canadian Medical Association Journal (insert article date, volume and issue numbers and page number(s)). © Canadian Medical Association (insert article year). This work is protected by copyright and the making of this copy was with the permission of the Canadian Medical Association Journal (www.cmaj.ca) and Access Copyright. Any alteration of its content or further copying in any form whatsoever is strictly prohibited unless otherwise permitted by law.

Kind regards,
Christina

Information Specialist
Access Copyright, The Canadian Copyright Licensing Agency
69 Yonge Street, Suite 1100
Toronto, ON M5E 1K3

From: Vlad Drobinin
Sent: Thursday, April 23, 2020 9:16 AM
To: Access Copyright Permissions Group
Subject: Inclusion of authored article in PhD Thesis

Figure B.3: License to include "Reliability of multimodal MRI brain measures in youth at risk for mental illness"



Publisher: John Wiley and Sons

© 2020 The Authors. Brain and Behavior published by Wiley Periodicals LLC

Open Access Article

This article is available under the terms of the Creative Commons Attribution License (CC BY) (which may be updated from time to time) and permits use, distribution and reproduction in any medium, provided that the Contribution is properly cited.

For an understanding of what is meant by the terms of the Creative Commons License, please refer to Wiley's Open Access Terms and Conditions.

Permission is not required for this type of reuse.

Wiley offers a professional reprint service for high quality reproduction of articles from over 1400 scientific and medical journals. Wiley's reprint service offers:

- Peer reviewed research or reviews
- Tailored collections of articles
- A professional high quality finish
- Glossy journal style color covers
- Company or brand customisation
- Language translations
- Prompt turnaround times and delivery directly to your office, warehouse or congress.

Please contact our Reprints department for a quotation. Email corporatesaleseurope@wiley.com or corporatesalesusa@wiley.com or corporatesalesDE@wiley.com.