



USING MACHINE LEARNING TO DISENTANGLE HETEROGENEITY WITHIN AND
BETWEEN PSYCHOSIS AND DEPRESSION: IMPROVING PATHWAYS FOR
PRECISION MEDICINE IN PSYCHIATRY

by

PARIS ALEXANDROS LALOUSIS

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Institute for Mental Health
Centre for Human Brain Health
School of Psychology
College of Life and Environmental Sciences
University of Birmingham
March 2022

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Dedication

This PhD thesis is dedicated to my dad (μπαμπά). Growing up you used to leave for work abroad very often and for long periods. I would miss you a lot and I would always wait for the Friday phone call where we would talk, and you would jokingly tell me that the statue of liberty, or whichever monument you were seeing from the ship you were working on, was saying hi to me.

When I became older our discussions became more centered around politics and football. My passion for both is a direct influence from you and that is something that will always stay with me. Six months into my PhD you suddenly left for your final trip. That night I received the worst phone call of my life. And every single day since then I have missed you and thought about you. I miss talking with you about what Olympiacos need to do in order to succeed in European football, what Manchester United need to do in order to reach the Sir Alex Ferguson heights again, and how British and Greek politics are interrelated in a unique way.

You left too early and without warning. I still remember your proud smile standing next to me during my MSc graduation and your elation when finding out that I got accepted into this PhD programme. It is still very hard to comprehend that you were not there for the majority of my PhD. I would give anything for a chance to be able to talk to you again one last time. You left too soon. This PhD thesis is dedicated to you μπαμπά μου. Μου λείπεις πολύ!

Acknowledgements

This PhD journey has been one filled with many experiences. During my PhD I unexpectedly lost a parent, had an anxiety relapse, went through a once in a lifetime pandemic, and a difficult breakup. During those times one of the things that kept me going was my passion for what I do and my dedication to effect change through my work. Despite life getting in the way of my PhD, I have absolutely loved every single minute of it and the last three and a half years of my life have been the richest and most fulfilling yet.

There are so many people I want to express my gratitude to, that it would take another PhD thesis to do them justice. First and foremost, I want to thank my amazing supervisory team. Professor Rachel Upthegrove has been a supervisor that one can dream of -her support (both professionally and personally) and belief in my strengths have been indispensable to my success. I wish everyone has a supervisor/manager who champions them the way Rachel has, and continues to, champion me. I still remember that Friday walk and talk from the Barberry to Muirhead tower where you explained the vision you had for my skillset!

Professor Stephen Wood was one of the reasons I came to Birmingham to do my MSc back in 2015. On an unseasonably hot April day, I walked to the fourth floor of the Frankland building to introduce myself and talk about my potential MSc project with him. Almost seven years later I am submitting my PhD which became possible thanks to him introducing me to the Birmingham psychosis research ecosystem.

In that Frankland room, I also had the chance to say hello to Doctor Renate Reniers who was then doing her postdoc. She would later become part of my PhD supervisory team and has always been there for me when needed -both academically and personally. Doctor Lianne Schmaal was also introduced to me by Stephen. She was the fourth addition to my supervisory team and despite never having met her in person her brilliance has been an important pillar to my PhD. I miss our (morning to me/evening to you) HYDRA meetings! I

want to thank all of you for your incredible scientific guidance but also for being compassionate human beings, providing the support and understanding I needed during critical times. Despite not officially being in my supervisory team, I also want to thank Professor Nikolaos Koutsouleris for encouraging me to challenge current scientific dogma with my research and for all his guidance and opportunities offered to me throughout this PhD. I also want to thank Doctors Jonathan Catling, Dominic Dwyer, Sian Lowri Griffiths, Markella Grigoriou, Peter Hansen, and Pavan Mallikarjun.

None of this would be feasible without my social support system of course. I want to express my deepest gratitude to my aunty Georgia Lalousis. Without her belief in me and her practical support I would never be where I am today. I want to thank my mum Sieglinde Eleonore Margarete Lalousis and my siblings Victoria, Lisa, and Nikos. You are all a source of comfort, love, and laughter and have played such an important part in the person I am today. The PhD gang –Imogen Mansfield, Georgia Bird, Timothy Foster, Fabian Burkhardt, Andrea Economou– I am indescribably lucky to have met such an incredible group of individuals. You have made the last three and a half years such an enjoyable ride! Your unconditional love and support have been instrumental to my happiness. Imogen and Georgia, you two have been particularly supportive of me in critical moments -I will never take your love and friendship for granted.

Daria Afanasyeva, Teodor Balint, Olivia Mitchell, and Sabrina Meechem, you are among the first friends I made when I came to Birmingham in 2015 and despite now living in different cities, I am grateful that our ties are as strong as they are. I cherish the times we have whenever we get the chance to meet up. Thank you for your continuous friendship and support. Francisca Krabberød, you were there for a big part of my PhD and I still remember reading my PhD acceptance with you. A big part of the contents of this PhD thesis was written while we explored life together. Your championing and pride for the work that I do

always humbled me and I appreciate you for that. Amoul Mangat, our camaraderie and bouldering and football related activities have offered me a great source of entertainment. I count myself lucky to have met such a great friend during my PhD.

Francesca Lewns, our venture into entrepreneurship and creation of Bloomwise has been equal parts challenging and rewarding. Our journey into the business world and the experiences we have shared in it have introduced an excitement that enriched my life in a unique way. Thank you for being an awesome business partner and friend! Marina Hoffman, your cut-throat, no-nonsense approach to life has always been an inspiration. Thank you for your friendship. Aggelos Adamidis, you have known me for 25(!) years, you probably need to be awarded a PhD for putting up with me for so long! Thank you for always being there buddy even though we now live in different countries. Anna Loli, your continuous care and support have been instrumental in difficult parts in my life. During the time of our friendship, I have also seen you flourish into the incredible woman you are today! I am lucky to be able to call you a friend. Bianca Diaconu and Aniela Wagstyl, your emotional support during a critical time in my life will never be forgotten.

Unfortunately, I cannot write out my gratitude for everyone, but you have also played an important role: Michalis Adamidis, Kleona Bezhani, Zoe Cox, Kaya Gromocki, Nierofer Hagos, Marianna Hurst, Angelos Iliopoulos, Panagiotis Koukouravas, Sofia Mamaki, Vasileios Mpakouros, Rebecca Walker, Ravi Punnett, and Shakira Smith. I want to thank all the students who I taught and supervised during my PhD. Your brilliance and passion for making a change are inspiring. I also want to thank all the amazing scientists who made the consortia, the data of which I used, possible, and all the patients who provided their data. Finally, I want to thank 14-year-old Paris for persevering.

-Amor Fati-

Role of Doctoral Researcher

The data used in chapters 4 and 5 comes from the PRONIA consortium. During the course of my PhD I was in charge of data storage and quality control for the PRONIA consortium in the Birmingham site. The data used in chapter 6 comes from the Australian Schizophrenia Research Bank (ASRB) dataset through external collaborations with Professor Cynthia Shannon Weickert. For all three empirical chapters (4,5, and 6) I a) conceptualised the research question, b) formed the hypotheses, c) developed analysis pipelines and/or scripts, d) performed the analyses, e) drafted the manuscripts, f) sent to my supervisory team for feedback, and g) then sent to all co-authors for comments.

Table of Contents	
Author's Declaration	i
Dedication	iv
Acknowledgements	v
Role of Doctoral Researcher	viii
Table of Contents	ix
Abstract.....	1
Chapter 1: Introduction	2
1.1 Schizophrenia	2
1.1.1 History of Psychosis	3
1.1.2 Aetiology of Psychosis	3
1.2 Depression.....	5
1.2.1 History of Depression	6
1.2.2 Aetiology of Depression	6
1.3 Treatment Challenges in Psychosis and Depression.....	8
1.4 Comorbidity of Psychosis and Depression	8
1.5 Diagnostic Approaches	11
1.5.1 Categorical Diagnoses	11
1.5.2 Dimensional Approach	13
1.5.3 Clinical Staging Model	13
1.5.4 The p factor	14
1.5.5. Network Theory of Mental Disorders.....	14
1.5.6 Transdiagnostic Approach	15
1.6 Machine Learning	15
1.6.1 Machine Learning Terms and Processes.....	16
1.6.2 Types of Machine Learning	17
1.6.3 Machine Learning's Potential	19
1.7 Conclusions.....	19
Chapter 2: Aims and Hypotheses	20
Chapter 3: Transdiagnostic Structural Neuroimaging features in Depression and Psychosis: A Systematic Review	22
3.1 Abstract	23
3.2 Introduction	24
3.3 Methods.....	25
3.3.1 Eligibility Criteria	25

3.3.2 Information Sources and Search Strategy	26
3.4 Results	28
3.4.1 Overview of Results	28
3.4.1.1 Gray Matter	28
3.4.1.2 White Matter	28
3.4.1.3 Risk of Bias	28
3.4.2 Gray Matter Volume	30
3.4.3 White Matter Volume.....	30
3.5 Discussion	32
3.5.1 Conclusions	35
Chapter 4: Heterogeneity and Classification of Recent Onset Psychosis and Depression: a Multimodal Machine Learning Approach	37
4.1 Abstract	39
4.2 Introduction	40
4.3 Materials and Methods	42
4.3.1 Study Design	42
4.3.2 Inclusion and Exclusion Criteria	43
4.3.3 Group Identification	43
4.3.4 MRI Imaging data acquisition, quality control, and preprocessing	44
4.3.5 Classification Models	44
4.3.6 Support Vector Machine Learning Analysis	45
4.3.7 Independent Validation	46
4.3.8 Supplementary Analyses	46
4.4 Results	46
4.4.1 Demographic Information	46
4.4.2 Machine Learning Analyses	47
4.4.2.1 Internal Validation of the Pure Group Differential Classifiers	47
4.4.2.1.1 Clinical and Neurocognitive Data	47
4.4.2.1.2 GMV Data	47
4.4.2.1.3 Stacking	48
4.4.2.2 Separability of Comorbid Groups	48
4.4.2.2.1 Clinical/Neurocognitive Data	48
4.4.2.2.2 GMV Data	48
4.4.2.2.3 Stacking	49
4.4.2.3 Independent Validation	49

4.4.3 Supplementary Analyses	49
4.5 Discussion	49
4.5.1 Strengths and Limitations.....	52
4.5.2 Conclusions	53
Chapter 5: Neurobiologically Based Stratification of Recent Onset Depression and Psychosis: Identification of Two Distinct Transdiagnostic Phenotypes	59
5.1 Abstract	62
5.2 Introduction	63
5.3 Methods	66
5.3.1 Study Design	66
5.3.2 Inclusion and Exclusion Criteria	66
5.3.3 MRI Imaging data acquisition, quality control, and preprocessing	66
5.3.4 Semi-Supervised Machine Learning Analysis	67
5.3.5 ComBat Harmonization.....	67
5.3.6 Model Training.....	67
5.3.7 Phenotype Characterization.....	68
5.3.8 Independent and External Validation	68
5.3.9 Predictive Utility	69
5.4 Results	69
5.4.1 Demographic Information	69
5.4.2 HYDRA Semi-Supervised Machine Learning Analysis	69
5.4.3 Cluster Statistical Significance.....	69
5.4.4 Clinical Characteristics Associated with Neuroanatomically Based Clusters	70
5.4.5 VBM Analysis of Neuroanatomically Based Clusters	71
5.4.6 Independent and External Validation	71
5.4.7 Prognostic Validation	72
5.5 Discussion	72
5.5.1 Strengths and Limitations.....	77
5.5.1 Conclusions	78
Chapter 6: Immune Subtypes of Schizophrenia and their Association with Brain Structure: a Semi-Supervised Machine Learning Examination of Inflammatory Heterogeneity.....	84
6.1 Research in context.....	86
6.2 Abstract.....	88
6.3 Introduction.....	90

6.3 Methods	92
6.3.1 Participants.....	92
6.3.2 Blood Based Biomarker Data	92
6.3.3 MRI Imaging data acquisition, quality control, and preprocessing	92
6.3.4 Neurocognition	93
6.3.5 External Dataset.....	93
6.3.6 Semi-Supervised Machine Learning Analysis.....	93
6.3.7 Model Training	94
6.3.8 Cluster Characterization.....	95
6.3.9 Cluster Solution Generalizability.....	95
6.3.10 Role of the funding source	95
6.4 Results	96
6.4.1 Demographic Information.....	96
6.4.2 HYDRA Semi-Supervised Machine Learning Analysis.....	96
6.4.3 Cluster Statistical Significance	96
6.4.4 Inflammatory Characteristics Analysis	97
6.4.5 VBM Analysis of Inflammatory Based Clusters	97
6.4.6 Neurocognitive Differences	97
6.4.7 External Dataset Cluster Exploration.....	98
6.5 Discussion.....	99
6.5.1 Strengths and Limitations.....	102
6.5.2 Conclusions.....	102
Chapter 7: Discussion	109
7.1 Summary of Chapter Findings	109
7.2 General Discussion.....	112
7.2.1 Depression and Psychosis	112
7.2.2 Heterogeneity and New Diagnostic Approaches	114
7.2.3 Biological Relevance and Precision Psychiatry	116
7.2.4 Clinical Applicability.....	118
7.3 Strengths and Limitations.....	120
7.4 Conclusions	120
References.....	122
Appendix A: Chapter 3 Supplement.....	170

Appendix B: Chapter 4 Supplement	174
Appendix C: Chapter 5 Supplement	213
Appendix D: Chapter 6 Supplement	264

Abstract

The aim of this PhD was to examine the clinical and biological -primarily structural brain- heterogeneity within and between depression and psychosis and provide tools for the improvement of diagnosis and targeted treatment. In chapter 3, a systematic review of structural neuroimaging studies in depression and psychosis identified potential transdiagnostic patterns of gray matter volume (GMV) and white matter volume (WMV) reductions in areas including the middle frontal gyrus, hippocampus, and left-sided posterior subgenual prefrontal cortex. In chapter 4, clinical/neurocognitive and neuroanatomical support vector machine (SVM) learning models demonstrated separability of prototypic depression from psychosis. Psychosis patients with affective comorbidity aligned more strongly to depressive rather than psychotic disease processes. In chapter 5, we identified two transdiagnostic neuroanatomically informed clusters which are clinically and biologically distinct, challenging current diagnostic boundaries in recent onset mental health disorders. In chapter 6, five clusters of schizophrenia with distinct immune signatures, associated with differing GMV and neurocognitive function were identified, with potential to inform the development of novel, targeted treatments. Overall, machine learning was utilised to elucidate and reduce heterogeneity within and between psychosis and depression, and identify biologically relevant and transdiagnostic subtypes that could become potential candidates for targeted treatment. The results are promising and challenge the current nosological system.

Chapter 1

Introduction

Mental health disorders affect over 1 billion people worldwide (roughly 16% of the world population). They are a major cause of disability and morbidity, accounting for five of the top 20 causes of global burden of disease (GBD) (Prince et al., 2007; Rehm & Shield, 2019). Depression and psychotic disorders constitute the majority of this burden, beginning in late adolescence/early adulthood and resulting in chronic disability for millions of people (Charlson et al., 2018; Vos et al., 2015; *WHO | Depression and Other Common Mental Disorders*, 2017).

Schizophrenia

Schizophrenia is one of the most common psychotic disorders. It is a severely debilitating disorder affecting approximately 1% of the world population, with onset usually during adolescence or early adulthood (Charlson et al., 2018; *The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992). Its core symptoms are delusions, which are fixed beliefs that are not amenable to change in light of conflicting evidence; hallucinations, which are sensory perceptions in the absence of a corresponding external or somatic stimulus; disorganized thinking; grossly disorganized or abnormal motor behaviour; and negative symptoms, which are functions normally present in healthy individuals but absent or subsided in affected ones and include blunted affect such as diminished facial and emotional expression, alogia (decrease in verbal output or verbal expressiveness), asociality (lack of involvement in social relationships of various kinds), avolition (a subjective reduction in interests, desires, and goals and a behavioural reduction of self-initiated and purposeful acts), and anhedonia (inability to experience pleasure from positive stimuli) (*The ICD-10 Classification of Mental and Behavioural Disorders : Clinical*

Descriptions and Diagnostic Guidelines, 1992). Despite its relatively low prevalence in the general population, schizophrenia contributes to 13.4 million years lived with disability globally and costs around US \$94-102 billion in the USA and €94 billion in Europe each year (Charlson et al., 2018; Chong et al., 2016; Olesen et al., 2012).

History of Psychosis

The concept of psychosis first appeared in 1841 by Canstatt as a shorthand for psychic neurosis, emphasizing the psychological manifestation of a disease of the brain.

Feuchtersleben has been credited with using the term psychosis for the first time as an alternative to mania and insanity and emphasizing the interplay between mental and physical processes in 1845 (Beer, 1995; Bürgy, 2008). It was not until 1891 that the concept of psychosis was narrowed down by Koch from meaning any disorder of the mind (as opposed to a disorder of the nervous system-neurosis) to mean “psychopathic inferiorities”, a term that was later renamed “abnormal personalities” by Schneider (Beer, 1995; Bürgy, 2008). The last decade of the 19th century brought about Kraepelin’s intricate work which culminated in 1899 with his sixth edition textbook with which he would influence the next 100 or so years of psychiatry by dichotomizing psychoses into manic depressive illnesses and dementia praecox the latter being renamed into schizophrenia by Bleuler in 1908 (Beer, 1995; Bürgy, 2008; Kendler, 2020).

Aetiology of Psychosis

The precise aetiology of psychotic disorders remains elusive despite decades of research (Radua et al., 2018). The age of onset of psychotic disorders however is well established in the mid-teens to early twenties (Kessler et al., 2007; McGorry et al., 2011). In terms of brain structure, the age of onset of illness is relevant- adolescence is a period during which the brain undergoes critical neural development and there is well established evidence for the

association between abnormal brain maturation and general psychopathology (Giedd et al., 2008; Gogtay et al., 2011). Indeed the neurodevelopmental hypothesis of schizophrenia posits the disorder stems from the interaction between brain “lesions” present from an early age and normal maturational processes during adolescence (Murray & Lewis, 1987; Weinberger, 1987). Normal gray matter volume (GMV) development follows an inverted “U” shape and reaches a peak in adolescence, and then a process of synaptic pruning begins. It is this synaptic pruning that has been suggested to be excessive in patients with schizophrenia due to early brain insults which disrupt normal brain maturational processes (Feinberg, 1982; Gogtay et al., 2011).

Several brain regions in schizophrenia have been found to have reduced GMV compared to healthy controls. Such regions include the thalamus, the caudate, the insula, the lateral prefrontal cortex, the anterior cingulate, the hippocampus, the superior temporal gyrus, the inferior frontal gyrus, and the cerebellum (Ellison-Wright et al., 2008; Fornito, Yücel, et al., 2009; Shenton et al., 2001). With regards to white matter, patients with schizophrenia have been found to have altered structural integrity in the cingulum bundles, the uncinate fasciculi, the internal capsules and the corpus callosum (Wheeler & Voineskos, 2014). It is not entirely clear whether these structural brain abnormalities are the cause or the effect of the disease or induced by interventions such as medication.

One model of explanation of brain abnormalities in schizophrenia and other psychoses is that of vulnerability-stress-inflammation. Stress is an additional risk in individuals who are genetically predisposed to the development of psychosis and who have also been exposed during an inflammatory response either during the second trimester of pregnancy or during later stages of central nervous system (CNS) development (Buka et al., 2008; Mednick et al., 1994; Müller et al., 2015). Higher production of the immune marker IL-1 β as well as higher

activation of microglia have been linked with GMV loss (Meisenzahl et al., 2001; Müller et al., 2015; Upthegrove & Barnes, 2014). Furthermore, elevated levels of IL-6, SERPINA3 mRNAs, and IL1 β have been found to be associated with greater GMV loss in patients with schizophrenia (Müller et al., 2015). Genomic predisposition and/or response to environmental factors leading to elevated pro-inflammatory cytokines may result in astrocytic dysfunction and/or microglial activation and subsequent dendritic pruning and synaptic changes thereby leading to GMV loss (Corsi-Zuelli & Deakin, 2021).

The aetiology of psychosis is not limited to brain structure and inflammation but includes many other processes (e.g., functional brain alterations, neurotransmitter signalling abnormalities, etc.) which are beyond the scope of this thesis.

Depression

Depression is a common, serious, recurrent, disorder with over 300 million people affected worldwide. It has a significant impact with high mortality rates, medical co-morbidity and a diminished role functioning and quality of life (*The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992; WHO | *Depression and Other Common Mental Disorders*, 2017). Its core characteristics are low mood, apathy, and social withdrawal with symptoms spanning the physiological, emotional, motivational, and cognitive domains present for 14 days and with a significant impact on functioning (Ménard et al., 2016; *The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992; WHO | *Depression and Other Common Mental Disorders*, 2017). The prevalence of depression is also growing, placing considerable strain on mental health services, with the cost of depression having been estimated to be around €118 billion in Europe and US\$71 billion in the USA (Olesen et al., 2012; Vigo et al., 2016; Vos et al., 2015). The prevalence of depression, just like psychosis

and the majority of mental health disorders, reaches its peak in late adolescence is established in adolescence and early adulthood, and a high proportion of young people with depression go on to have recurrent episodes or fail to recover fully (*The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992; WHO | *Depression and Other Common Mental Disorders*, 2017).

History of Depression

The concept of depression can be traced back to ancient Egypt where descriptions of a depressive state appear in papyri and ancient Greece where Hippocrates is credited with making the first hypothesis that depression is a dysfunction of the brain due to an excess of black bile or *melaina chole* in Greek which gave depression its first name (*Melancholia*) (Davison, 2006). It was not until the early 19th century that *melancholia* became a dissatisfactory term and was attempted to be replaced by numerous new terms with varying success. Griesinger attempted to unify all those terms and other forms of “insanity” under the term psychosis in his *unitary psychosis hypothesis* (Davison, 2006). Kraepelin would then make his seminal distinction between *dementia praecox* and manic-depressive illness as detailed above. After almost a century of debate surrounding endogenous, exogenous, and reactive descriptions of depression, the Diagnostic and Statistical Manual of Mental Disorders (DSM)-III introduced the term and concept of Major Depressive Disorder which has been widely, although not universally, accepted (American Psychiatric Association, 1980).

Aetiology of Depression

Similarly to psychosis, the precise aetiology of depression remains unclear despite decades of research (Brand et al., 2015; Dean & Keshavan, 2017). In terms of brain structure, lower cortical thickness and cell densities in the prefrontal cortex (PFC) has been one of the most

consistent findings in histopathological studies of depression (Ménard et al., 2016; Rajkowska et al., 1999). Moreover, patients with depression have been found to have decreased GMV in the hippocampus, the anterior cingulate cortex, and the orbitofrontal cortex (Brand et al., 2015; Campbell & MacQueen, 2004; Schmaal et al., 2016). Overall structural dysconnectivity has been suggested in patients with depression (van Velzen et al., 2020). Fractional anisotropy (FA) reductions in the corpus callosum, the anterior limb of the internal capsule, the right inferior temporal gyrus, and the right superior frontal gyrus which are associated with symptom severity and duration of illness have been found in patients with depression in comparison with healthy controls (Chen et al., 2017).

The pathophysiology of depression may be explained, in part, by inflammation. Patients with depression have been found to have higher levels of proinflammatory cytokines, chemokines, cellular adhesion molecules and acute phase proteins which interact with neurotransmitter metabolism, neural plasticity and behaviour, and neuroendocrine function (A. H. Miller et al., 2009; Osimo et al., 2019; Raison et al., 2006). A potential explanation of the relationship between low-grade inflammation and depression is that of social signal transduction theory of depression. This hypothesis postulates that the immune system's reaction to experiences of social threat and adversity -such as modern-day social, symbolic, and imagined threats- is to up-regulate its inflammatory components. This up-regulation can in turn bring about changes in brain morphology and behaviour that signal the initiation of depressive symptoms (Slavich & Irwin, 2014). Another hypothesis is that inflammation in depression is due to chronic stress or unhealthy diet/lifestyle, and microbiome abnormalities among other factors (Milaneschi et al., 2020, 2021; Toenders et al., 2021)

Similarly to psychosis, the aetiology of depression is not limited to brain structure and inflammation but includes many other processes which are beyond the scope of this thesis.

Treatment challenges in psychosis and depression

Both psychosis and depression present clinicians with treatment challenges. Only 14-20% of people with psychosis will have a full recovery, with the majority having recurrent episodes and relapses. Three quarters of people who are diagnosed with schizophrenia experience recurrent relapse and continued disability (S. Brown et al., 2010; Health (UK), 2014; Owen et al., 2016). Initial treatment response is positive for the majority of patients, but it has been estimated that around 80% of people with psychosis relapse within five years of a first episode that has been treated (S. Brown et al., 2010). Similarly, only 25% of people with depression achieve full remission and response to pharmacological treatment, with the remainder achieving partial response or response without remission (Kern et al., 2020). More importantly whilst psychological treatments and medication can be partially effective in people with depression and people with psychosis, presently it is impossible to accurately predict who will need longer-term intervention or monitoring, or which treatment course is best for each individual.

Comorbidity of Psychosis and Depression

The comorbidity between schizophrenia and depression is exceptionally high. In schizophrenia, there is a reported prevalence of depression in around 40% of people. In acute episodes of schizophrenia, the prevalence of depression is higher still at around 60% (Conley et al., 2007; Upthegrove et al., 2017). Depression is the most significant risk factor for suicidal behaviour in first-episode psychosis (FEP) and also has long-term consequences for other poor outcomes, including social and occupational recovery, an increased risk of substance misuse and reduced medication adherence and reduced quality of life (Conley et al., 2007; Koreen et al., 1993; WHO | *Depression and Other Common Mental Disorders*, 2017). Conversely, psychotic symptoms occur in 14-19% of people diagnosed with major

depressive disorder and there exists a positive correlation between the proportion of reported hallucinations and/or delusions and the number of existent depressive symptoms (Johnson et al., 1991; Ohayon & Schatzberg, 2002). Furthermore, the presence of psychotic features in depression contribute to a more severe outlook, as reflected in an increased risk of suicide attempts, hospitalization, relapse, persistence of symptoms over one year, and financial dependency compared to depression with no psychotic features (Ohayon & Schatzberg, 2002).

Despite the known association between schizophrenia, or more broadly defined psychotic symptoms, depression, and poor outcomes, the diagnosis and management of comorbidity presents a problem for clinicians, with a reported rate of misdiagnosis of over 27% (Rothschild et al., 2008). Currently, it is theorized that three potentially different pathways explain the comorbidity of depression in psychotic disorders: depression as a reaction to the diagnosis and experience of psychosis, depression as evidence of common historical childhood adversity trauma, and depression which is intrinsic to psychosis (Birchwood et al., 2005). A number of studies also suggest that depression can increase before the onset of first-episode psychosis or psychotic relapse, suggesting that depressive symptoms rather than ‘comorbidities’ may be more causally related to the development of psychosis (Conley et al., 2007; WHO | *Depression and Other Common Mental Disorders*, 2017). Depression could potentially provide an aetiological pathway across mood and psychotic disorders, borrowing from Foulds’s hierarchy of classes of personal illness (Foulds & Bedford, 1975; Upthegrove et al., 2017). Both latter two models, that of depression as intrinsic or seen within a hierarchical structure, suggest that in schizophrenia, depression is of secondary order in nosological status, whilst in primary major depressive disorder, psychotic symptoms are secondary.

Regarding the presence of psychosis in Major Depressive Disorder, the two main models of explanation are the severity model and the vulnerability model (Tonna et al., 2012). The severity model posits that the presence of psychotic symptoms in depression are indicators of severe depression whereas the vulnerability model suggests that a psychotic presentation of depression is predisposed by personality characteristics (Harrow et al., 2000; Tonna et al., 2012). A major distinction is made between mood-congruent and mood-incongruent psychotic symptoms in depression. Mood-congruent psychotic symptoms are symptoms that are congruent with the theme of depression (guilt, low mood, etc.) while mood-incongruent psychotic symptoms are symptoms that are not congruent with the depressive theme (Kendler, 1991; Tonna et al., 2012). Mood-incongruent psychotic symptoms have been associated with poorer outcome and more severe depression (Harrow et al., 2000; Tonna et al., 2012).

However, whilst psychosis is characterised by positive symptoms such as delusions and hallucinations, and depression by low mood, withdrawal and apathy (*The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992), certain features are seen in both disorders, for example loss of motivation, poor social functioning, and anhedonia (Krynicky et al., 2018). Anhedonia may be of particular interest with regards to its transdiagnostic qualities, with certain subtypes such as motivational and consummatory anhedonia being ascribed as more specifically to psychosis and depression respectively (Upthegrove et al., 2017). Furthermore, it has been shown that both disorders are associated with working memory, executive functioning, and verbal fluency deficits, with previous studies failing to find a distinct neurocognitive profile between the two disorders (Moritz et al., 2002).

Depression and psychosis share certain characteristics identified in previous structural magnetic resonance imaging (sMRI) research. Hippocampal GMV reductions are observed in both disorders (Meisenzahl et al., 2010; Upthegrove et al., 2017). Symptoms of depression such as withdrawal, blunted affect, and alogia, are associated with GMV reduction in the cerebellum, while anhedonia and avolition are negatively correlated with left anterior limb of internal capsule white-matter volume (WMV) and positively correlated with left superior longitudinal fasciculus WMV (Chuang et al., 2014). Moreover, in functional MRI studies (fMRI) both disorders have been associated with heightened brain activation in response to fearful facial expressions specifically in the thalamus as well as with regions such as the insula, the amygdala, the prefrontal cortex, and the hippocampus, which are central to emotional processing (Broome et al., 2015; Busatto, 2013; Kumari et al., 2016).

Diagnostic Approaches

Categorical Diagnoses

Currently, mental health disorders in general, and psychosis and depression specifically are understood and defined in a categorical framework. The International Classification of Diseases (ICD) and the DSM which are the most widely used classification systems in psychiatry represent a categorical understanding of mental health disorders which views them as discrete non-overlapping entities (American Psychiatric Association & American Psychiatric Association, 2013; *The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992). A categorical approach makes sense in the world of clinical medicine due to the inherent dichotomous nature of the decisions that have to be made by clinicians such as to treat or not to treat, to discharge or to hospitalize etc. (Kendler, 2018). These diagnoses are based on a descriptivist approach that uses signs and symptoms to assign people to categories -and while they have high reliability,

their usefulness is questionable (Cuthbert, 2014; First et al., 2018; Fusar-Poli, Solmi, et al., 2019). Moreover, comorbidity is a great challenge that diagnostic systems face. Comorbidity rates in psychiatry are very high and follow a rule of 50% where half of people who meet diagnostic criteria for one mental health disorder also meet diagnostic criteria for a separate disorder at the same time, half of people who meet diagnostic criteria for two disorders at the same time also meet diagnostic criteria for a third disorder and so on (Newman et al., 1998). Perhaps due to the lack of accepted biological or genetic markers, diagnostic categories have remained in place for over 100 years. Specifically in depression and psychosis, diagnoses have survived since Kraepelin's dichotomy due to the fact that some individuals do indeed fit within these nosological entities (Potuzak et al., 2012). Psychosis is classified into two main categories: non-affective (schizophrenia, schizophreniform disorder, brief psychotic disorder, delusional disorder, and psychotic disorder not otherwise specified) and affective psychosis (bipolar disorder, schizoaffective disorder, and depression with psychotic features) (American Psychiatric Association & American Psychiatric Association, 2013; *The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992).

According to DSM-5 criteria an estimated 16,400 combinations of symptoms and sub-symptoms can lead to the diagnosis of a depressive episode with 1,030 unique symptom profiles having been identified. Depression is categorized into two main categories: major depressive disorder (MDD) and MDD with psychotic features (with positive symptoms being mood congruent and secondary to mood symptoms) (American Psychiatric Association & American Psychiatric Association, 2013; Ballard et al., 2019; Fried & Nesse, 2015).

Dimensional Approach

A dimensional approach has been suggested as an alternative to categorical diagnoses. This approach posits that disorders such as psychosis and depression arise from pluripotential states in which different groups of correlated symptom dimensions a) might coexist in the same individual, b) are not exclusive but complementary, c) cut across all types of diagnoses. This approach also theorizes that symptoms exist along a continuum in the general population and that mental health disorders might represent the extreme end of the distribution of a combination of symptoms (Hengartner & Lehmann, 2017; Kotov et al., 2017; McGorry, 2013; Peralta & Cuesta, 2003; Van Os et al., 1999).

Clinical Staging Model

A clinical staging approach hypothesizes that mental health disorders occur in five stages with the first stage being the “at-risk” asymptomatic stage and the second stage being the stage during which individuals have undifferentiated, general symptoms which then worsen and show “syndromal specificity”. In the third stage individuals experience a first episode of a full threshold syndrome which could be followed by persistence or recurrence in their symptoms (fourth stage). The fifth stage is characterized by severe and unremitting illness. Although remission and amelioration are less likely with each advancing stage, they are still possible throughout. One of the central tenets of this approach is agnosticism towards traditional symptom-based categorical diagnostic boundaries and adoption of neurobiologically-informed approaches. Furthermore, it is suggested that interventions should be stage-specific. It is also recommended that pathophysiology is distinguished from etiological factors. Finally it is suggested that new symptom clusters could inform evolving syndromal coherence to improve categorical representation (McGorry et al., 2014).

The p factor

The p factor is another approach that attempts to transcend traditional categorical nosologies. This approach suggests the structure of psychopathology can be summarized by three dimensions: an internalizing liability to depression and anxiety; an externalizing liability to antisocial and substance-used disorders; and a thought disorder liability to symptoms of psychosis. Furthermore it posits that there is a general underlying dimension that characterizes an individual's propensity to develop any and all forms of psychopathology (the p factor) (Caspi et al., 2014). While a factor modelling approach is useful when measuring constructs that are not directly observable there are risks in the causal interpretations of factors as entities (van Bork et al., 2017).

Network Theory of Mental Disorders

A network theory of mental disorders takes issue with the terms of mental disease and mental disorder. The use of the term disease implies that there is an underlying clear aetiology and common pathological pathway from which symptoms arise. The use of the term mental disorder on the other hand implies a syndromic constellation of symptoms. The fact that other than a few constellations of symptoms in mental disorders have been proven, common causes create a problem that is unique to psychiatry. In other areas of medicine removing the disease entity (for example a tumour) removes the common cause of phenomenologically distinct symptoms because of the underlying causal homogeneity. Due to the fact that central disease mechanisms have not been identified in mental disorders this strategy is not effective or even feasible in psychiatry. The network theory of mental disorders suggests that central mental disorder mechanisms have not been identified because they do not exist. Rather, psychiatric symptoms could potentially cause each other instead of emanating from a common cause. This interaction between symptoms can be understood as a network in which symptoms are

nodes and the interaction between them are connections between nodes. More pertinently, this theory posits that symptoms do not inherently “belong” in any disorder, rather they exist transdiagnostically and it is specific interactions that give rise to what we understand as different psychiatric diagnoses (Borsboom, 2017; de Boer et al., 2021).

Transdiagnostic Approach

A transdiagnostic approach theorizes that symptoms exist along a continuum but also radically disregards diagnoses and aspires to replace them with new entities based on a data-driven bottom-up approach. Transdiagnostic approaches suggest that disorders share common aetiological and maintenance processes and that disorder-specific treatments overlook comorbidity (Fusar-Poli, Solmi, et al., 2019). The Research Domain Criteria project (RDoc) is an example of the transdiagnostic approach. Its aims are to a) bring together clinical and basic science in order to identify fundamental behavioural components that might transcend categorical diagnoses and are based on neuroscience, b) determine the variation from normal to abnormal within these fundamental components, c) develop reliable and valid measures of these components which will then be used in research and clinical settings, and d) to integrate the behavioural, genetic, neurobiological, environmental, and experiential domains that comprise mental disorders (Cuthbert, 2014; Fusar-Poli, Solmi, et al., 2019).

Machine learning

The ultimate goal in psychiatry and clinical psychology is to have an accurate view of mental health disorders with a homogeneous underlying pathophysiology and predictable responses to effective treatment. Research in psychiatry has encountered difficulties in the translation from neuroscience into clinical practice with findings often failing to inform new treatments (Bzdok & Meyer-Lindenberg, 2018). This has been postulated to be the result of traditional statistics seeking to reject the null hypothesis by considering the entirety of a data sample

rather than focussing on individual cases (Bzdok & Meyer-Lindenberg, 2018; Dwyer et al., 2018).

Machine Learning Terms and processes

Machine learning is a branch of artificial intelligence that can be broadly categorized into supervised machine learning (SML), unsupervised machine learning (UML), semi-supervised learning, clustering, neural-networks, representation learning, and reinforcement learning. In machine learning the dataset is often split into two parts. The training data and the test data. Training data are the data on which the mathematical models used by the machine learning algorithms are applied to (“trained on”) in order to make prediction or decisions without being explicitly programmed to do so. Test data on the other hand are the new, independent data on which the generated model is applied to in order to evaluate the performance of the trained model (Bishop, 2006; Bzdok & Meyer-Lindenberg, 2018). Model validation takes place by means of cross-validation. Cross-validation is used in order to estimate how accurate a predictive or decision model is. Models are trained on the training data whose labels and outcomes are already known and then applied on new test data whose labels and outcomes are unknown. The accuracy of a model is evaluated by how successful predictions or decisions it makes on the test data (Bishop, 2006).

There are many different types of cross-validation (e.g. leave-p-out cross-validation, leave-one-out cross-validation, k-fold cross-validation etc.), however the focus in this thesis will be repeated nested cross-validation- which is the gold standard in the field (Filzmoser et al., 2009). While cross-validation is good practice in machine learning, it has two main weaknesses. Splitting the whole dataset into training and test data might make either set unrepresentative of the whole, hence the produced model might end up being skewed. Secondly, when evaluating different models with different sets of features during cross-

validation the models are tested on the test data. After evaluating the first model, the test data have already been accessed and therefore are no longer new and independent. Any subsequent models will be evaluated on data that have already been seen and the winning model will be one that performs the best on that data and no longer on new and independent data like one would encounter in a real-world scenario. This is a type of “data leakage” (Bishop, 2006).

Repeated nested cross-validation deals with this issue by splitting the whole data into two main cross-validation cycles, the outer cross-validation cycle and the inner cross-validation cycle. Models are trained in the inner-cross validation cycle and then applied to the data in the outer cross-validation cycle. Each cycle is divided into k folds (usually 10) each containing independent data. Furthermore, in each cycle one-fold is held out for testing. The data gets trained in each of the inner-cross validation folds in the inner cycle and then applied to the inner cross-validation held out fold. The data in the inner cross-validation gets shuffled around so that the folds contain different data and the process gets repeated until all folds have been trained and produced a model and then permuted k times (usually 10). The models that are produced in the inner cross-validation cycle are then applied to all the folds apart from the one held out in the outer cross-validation cycle and permuted k times (usually 10). The models that have the best performances are then applied to the held out fold and permuted k times (usually 10) (Bishop, 2006; Filzmoser et al., 2009).

Types of machine learning

Supervised machine learning is a type of machine learning whose objective is to learn a function that can map an input to an output based on prior knowledge of the types of inputs and outputs. Data have the form of features that describe a problem as well as a target variable (e.g. group membership) (Bishop, 2006; Bzdok & Meyer-Lindenberg, 2018; Russell

& Norvig, 1995). Supervised machine learning has many advantages such as a mechanistic insight into its processes, good knowledge of the input data, and theoretical input into class determination. Some of its disadvantages include a tendency to overfit the data, increased computational time, and the need to have prior knowledge of the classes you want to predict.

In unsupervised machine learning, algorithms are used in order to discover hidden patterns or data grouping without prior knowledge or labelling from humans. These algorithms use a certain criterion to partition sets of datapoints into a number of groups according to their similarity (Bishop, 2006; Bzdok & Meyer-Lindenberg, 2018). Its advantages are that there is no need for prior knowledge of the classes in the dataset and therefore provides opportunities to uncover hidden patterns. Its disadvantages are that there is usually lack of mechanistic insight into the algorithm's processes and there is no "ground truth" against which its results can be measured and also the fact that purely unsupervised machine learning models might reflect underlying nuisance variance (Dinga et al., 2019).

Semi-supervised machine learning is a hybrid method combining methods from both supervised and unsupervised methods. In semi-supervised approaches algorithms are used in order to discover data groupings in unlabelled data for which the ground truth is not known but the algorithm is also provided with labels for some of the cases (Chapelle et al., 2010). Semi-supervised machine learning models have the advantage of exquisitely dissecting the heterogeneity in datasets by portioning the unlabelled data based on the patterns or transformations between the sub-populations (i.e., clusters) from the unlabelled data and the labelled data (reference group). Furthermore, these models typically regress out covariates of no interest. Semi-supervised machine learning has a similar disadvantage to unsupervised machine learning in that mechanistic insight is not always easy.

Machine Learning's Potential

Machine learning holds significant potential in solving the difficulties faced by traditional statistics in mental health research (Bzdok & Meyer-Lindenberg, 2018). While inferential classical statistics rely on making certain assumptions about the data (such as the normality of the data's distribution, homogeneity of covariance, independence of residuals etc.), machine learning methods rely on few assumptions, allow the data to “speak for themselves”, and allow for the mining of structured knowledge from extensive data that can then be applied to individual cases, allowing for personalised approaches in mental health diagnosis and treatment (Bzdok & Meyer-Lindenberg, 2018; Chen & Zhu, 2001).

Conclusions

Mental illness has an intricate history that has led to the current nosological system which presents diagnostic, prognostic, and treatment challenges to clinicians. Aetiological pathways are not clear and heterogeneity prevents targeted approaches and perpetuates the lack of biomedical translation from neuroscience to clinical care. Harnessing well characterised multi-site datasets and utilising robust and novel machine learning frameworks can lead to improved pathways for precision medicine in psychiatry which is the aim of this PhD thesis.

Chapter 2

Aims and hypotheses

The overarching aim of the studies contained in this thesis is to elucidate the clinical and biological -primarily structural brain- heterogeneity within and between depression and psychosis and provide tools for the improvement of diagnosis and targeted treatment.

In chapter 3, I present a systematic review of structural neuroimaging studies in depression and psychosis, aimed to establish the current evidence of neuroanatomical similarities and differences between psychosis and depression. This paper is in submission to *Neuroimage: Clinical*.

In chapter 4, the diagnostic heterogeneity within and across psychosis and depression is investigated with common, comorbid transdiagnostic phenotypic, neuroanatomical, and neurocognitive features to establish the disease processes of comorbid illness. We hypothesized that developed models would correctly classify diagnostic groups without comorbid symptoms, in keeping with evidence of the utility in categorical diagnostic structure, and that grey matter volume (GMV) would add classification accuracy. We further hypothesized that a reduction of classification accuracy would be seen in groups with comorbid symptoms. This paper is published in *Schizophrenia Bulletin*.

In chapter 5, I present a neurobiologically-based diagnostic structures were formed using neuroanatomical data across depression and psychosis and their prognostic utility was investigated.

This paper has been published in *Biological Psychiatry*.

In chapter 6, I investigate the heterogeneity of immune dysfunction and GMV in schizophrenia to elucidate its aetiological pathways to inflammation and identify potentially

clinically relevant targets for novel treatments. This paper is under review by the *American Journal of Psychiatry*.

Finally, in chapter 7, the implications of the results of the studies contained in this thesis are discussed and their potential in informing the direction of precision psychiatry is evaluated.

Chapter 3

Transdiagnostic Structural Neuroimaging features in Depression and Psychosis: A Systematic Review

Paris Alexandros Lalousis, MSc, Stephen Wood, PhD, Renater Reniers, PhD, Lianne Schmaal, PhD, Hannah Azam, MSc, Antonella Mazziota, MSc, Hasson Saeed, Msc, Charlotte Wragg, MSc, Rachel Upthegrove, MBBS FRCPsych, PhD

Abstract

Background: Previous research suggests that there may be similarities in structural brain changes seen in patients with depression and psychosis compared to healthy controls.

However, there is yet no systematic review collating studies comparing structural brain changes in depression and psychosis. Establishing shared and specific neuroanatomical features could aid the investigation of underlying biological processes.

Aims: To identify structural neuroimaging similarities and differences between patients with depression and psychosis.

Method: We searched PubMed, PsychInfo, Embase, NICE Evidence, Medline and the Cochrane Library were searched from inception to 30/06/2021 using relevant subject headings (controlled vocabularies) and search syntax. Papers were assessed for quality using the Newcastle-Ottawa Scale.

Results: Five hundred and twenty papers were retrieved, seven met inclusion criteria. In narrative collation of results, grey matter volume (GMV) reductions were found in the medial frontal gyrus (MFG), hippocampus and left-sided posterior subgenual prefrontal cortex in both psychosis and depression. GMV reductions affected more brain regions in psychosis, including in the insula and thalamus. White matter volume (WMV) decline was found in both depression and psychosis. Reduced fractional anisotropy (FA) was more commonly seen in depression.

Conclusions: Our results suggest potential transdiagnostic patterns of GMV and WMV reductions in areas including the MFG, hippocampus, and left-sided posterior subgenual prefrontal cortex. These could be investigated as a future biomarker of transdiagnostic signature across mental illnesses. However, due to the limited number and poor quality of studies future research in large samples and harmonised imaging data is first needed.

Introduction

Depression and psychotic disorders constitute the major challenges of mental ill health to the world's population (Charlson et al., 2018; Vos et al., 2015; *WHO | Depression and Other Common Mental Disorders*, 2017). Major Depressive Disorder (MDD) has a lifetime prevalence rate of approximately 20% and is one of the most common and disabling psychiatric diseases globally (*WHO | Depression and Other Common Mental Disorders*, 2017). Psychotic disorders, including schizophrenia affect up to 3% of the world population (Charlson et al., 2018; *The ICD-10 Classification of Mental and Behavioural Disorders : Clinical Descriptions and Diagnostic Guidelines*, 1992) and are severely debilitating.

Despite depression and psychosis being two different diagnostic structures, they may share some aetiologies and maintenance processes. Depression in schizophrenia has long been a taxonomic challenge, with various approaches to address it including that core, deficit schizophrenia is “non-affective” and early more heterogeneous stages of psychosis have higher affective load (Upthegrove et al., 2017). Craddock and Owen (2010) argue genetic studies reinforce evidence that suggests a clear dichotomy between mood and psychotic disorders is absent.

Symptoms such as social withdrawal, blunted affect, anhedonia, avolition, and alogia are observed both in psychosis and depression (Krynicky et al., 2018). Social withdrawal, blunted affect, and alogia are associated with GMV reduction in the cerebellum, while anhedonia and avolition are negatively correlated with left anterior limb of internal capsule white-matter volume (WMV) and positively correlated with left superior longitudinal fasciculus WMV (Chuang et al., 2014; Meisenzahl et al., 2010). Structural magnetic resonance imaging (sMRI) has shown that patients with psychiatric disorders may have distinct brain changes compared to healthy controls. Depression and psychosis may share certain neuroanatomical

characteristics; e.g. hippocampal grey matter volume (GMV) reductions are observed in both disorders (Chuang et al., 2014; Meisenzahl et al., 2010). Overlap in neuroanatomical characteristics between the two disorders may be suggestive of common aetiological mechanisms, with depression potentially playing an intrinsic role in the pathogenesis of schizophrenia (Birchwood et al., 2005; Sönmez et al., 2013). However, to our knowledge there has been no previous systematic review of structural neuroimaging studies comparing depression to psychosis, and here we aim to address this gap.

Methods

The protocol was pre-registered on PROSPERO ID CRD42020190200 and is reported in line with Preferred Reporting lines for Systematic review and Meta-Analyses (PRISMA) guideline (Moher, 2009).

Eligibility Criteria

Eligibility criteria were established following PICOS format - patients/population, intervention, comparisons and outcomes (Liberati et al., 2009): a) only human participants; b) ages of 14-70; c) include participants with a diagnosis of major depressive disorder or recent onset depression AND patients with a diagnosis of first episode psychosis, or recent onset psychosis, or schizophrenia spectrum disorders, or schizophrenia, or schizophreniform disorders; d) studies reporting primary cross-sectional or longitudinal or follow up data; e) studies reporting GMV, or cortical thickness, or gyrification, or surface area, or white matter anisotropy or diffusion tensor imaging.

Studies including only bipolar disorder were excluded however, if these also included psychosis and depression, they were included as it may be possible to extract data from the relevant groups of interest. Animal studies, non-English language studies and pharmacological interventions were excluded.

Two authors (H.A. and C.W) screened titles and abstracts and then assessed eligible studies for potential inclusion with a thorough full text search. Two separate authors (A.M. and H.S.) checked the final list and any conflicts between authors were discussed and consensus achieved.

Information Sources and Search Strategy

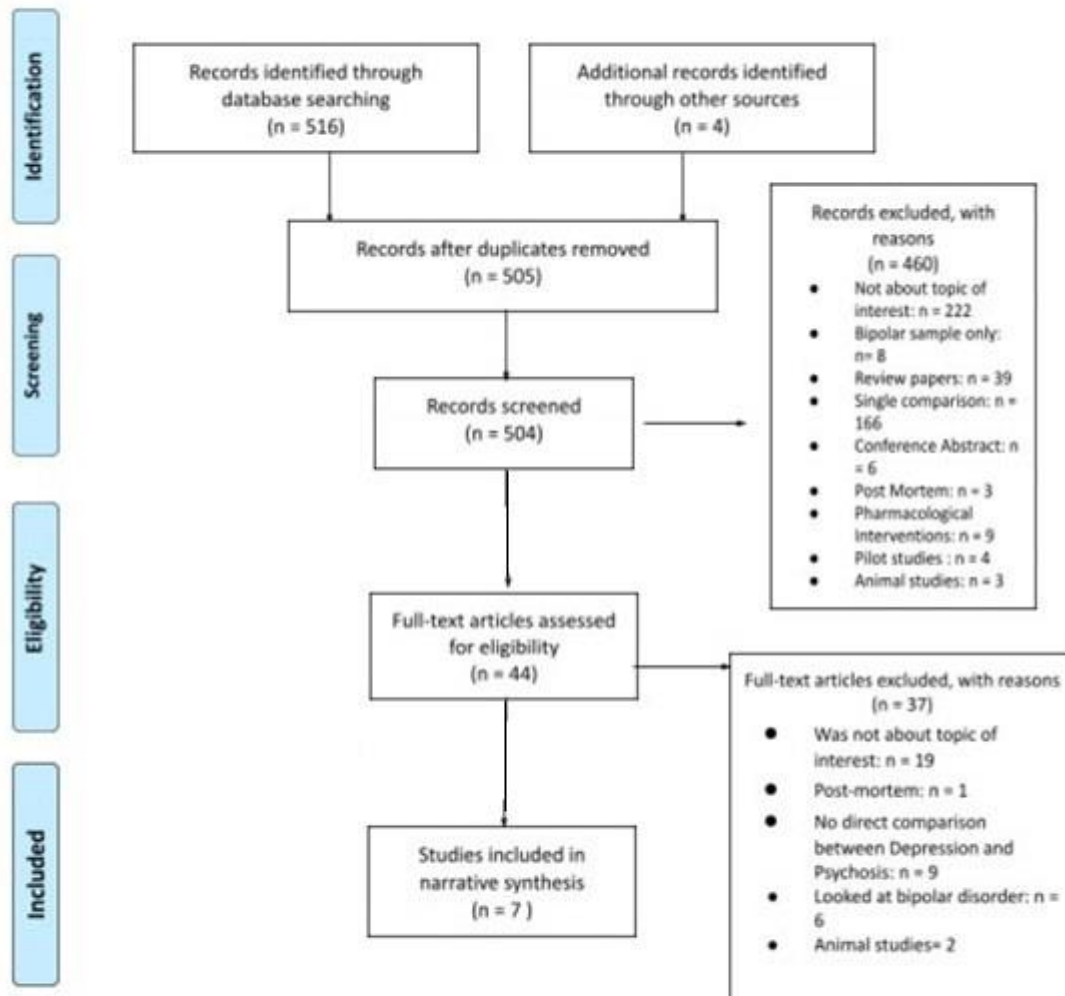
PubMed, PsychInfo, Embase, NICE Evidence, Medline and the Cochrane Library were searched from inception to 30/06/2021 using relevant subject headings (controlled vocabularies) and search syntax. See supplementary material and PROSPERO registration for full search terms.

A systematic search combining terms describing psychosis with AND to terms describing depression and with OR to terms describing different structural neuroimaging modalities was conducted on the following databases: PubMed, PsychInfo, Embase, NICE Evidence, Medline and the Cochrane Library from inception to 30/06/2020. The search terms were adjusted for different databases in subject headings and proximity operators. Please see supplement 1.1 for full details on the search strategy.

Reference lists from included studies were also hand searched for studies which met eligibility criteria that may not have been found from the databases.



PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 1. PRISMA flow diagram documenting the study selection process including the identification, screening, eligibility and included studies stages.

Results

The systematic search identified 516 unique abstracts. Full text assessments was performed on 44 studies, with 37 of these excluded. Seven studies reporting on seven unique samples with a total of 1,452 participants fulfilled eligibility criteria were included in the review. The mean age of depression patients in the sample was 36.0 (12.1). The mean age of psychosis patients was 31.9 (8.8). Five studies investigated gray matter and four studies explored white matter.

Overview of results

Grey Matter: Five studies (Coryell et al., 2005; Y. Jiang et al., 2019; Keller et al., 2008; Meisenzahl et al., 2010; Ota et al., 2013) identified: GMV reduction in the bilateral middle frontal gyrus was reported in both depression and psychosis; hippocampal GMV reductions both in depression and psychosis; bilateral amygdala GMV reduction in psychotic depression compared to non-psychotic depression, and reduced GMV in the insula and the thalamus in psychosis compared to depression.

White Matter: Three studies (Cui et al., 2020; Kochunov et al., 2013; Ota et al., 2013) reported reduced WMV volume fractional anisotropy (FA) in the corpus callosum and bilateral anterior and posterior insula in depression compared to psychosis.

Risk of bias

The quality check was completed using the Newcastle-Ottawa Scale (NOS), (Wells et al., 2000). Four out of the seven studies were rated 4 stars which is considered low quality. See supplementary table S1 for full details.

Gray matter volume

Jiang et al. (2019), conducted voxel-based-morphometric analysis of the total brain volume between patients with schizophrenia and major depressive disorder (MDD). Results found four clusters with significant group effects. In detail, lower GMV in the bilateral MFG, the medial prefrontal cortex (MPFC), and the rectus gyrus was found in patients with schizophrenia compared to patients with MDD. The authors investigated the correlation and causal effects between GMV alterations and clinical variables for each patient group. Specifically, in the schizophrenia group reduction in GMV was negatively correlated with PANSS negative scores. Using Granger Causality Analysis (GCA) they found that the causal link between the right cerebellum and the medial prefrontal cortex had a positive correlation with PANSS negative scores. Finally, authors found a significant negative correlation between GMV in the MPFC and the causal link between the right cerebellum and the MPFC. However, this was not the same for the depression group where no significant correlations or causal effects between GMV alterations and clinical features were observed.

Coryell et al. (2005) looked at GMV in the anterior and posterior Subgenual Prefrontal Cortex (SGPFC) in patients with psychotic depression compared to patients with schizophrenia and healthy controls. Results found that the SGPFC GMV did not significantly differ between patients with psychotic depression and schizophrenia. However, patients with psychotic depression reported the smallest volumes in the left-sided posterior SGPFC. The differences found in the left-sided posterior SGPFC volumes were compared with other structure measures within this region including cortical depth and surface area. Surface area did not differ significantly between groups; however, the depth of the surface area was significantly less in the psychotic depression group compared to the schizophrenia group. On

follow up, increases in the posterior SGPFC volume were found in the psychotic depression group only.

Keller et al. (2008) performed a region of interest (ROI) analysis to investigate volumetric differences in amygdala and hippocampus in patients with psychotic depression and nonpsychotic depression. Patients with psychotic depression had marginally smaller bilateral amygdala volumes than patients with non-psychotic depression. The study found no significant differences in bilateral hippocampal volume between groups. However, researchers report that their sample of participants did not differ on measures of recurrent chronic depression, which might be the reason for no differences in hippocampal volume.

Meisenzahl et al. (2010) looked at differences in hippocampal volume between patients with depression and schizophrenia and found a significant effect of diagnosis on reduced bilateral hippocampal GMV and hippocampal WMV. Furthermore, when studying differences between patients with first episode of depression and recurrently ill patients with depression the authors reported found no significant differences concerning hippocampal volume measurements.

Ota et al. (2013) found that patients with schizophrenia showed reduced GMV in the bilateral anterior and right posterior insula and the bilateral anterolateral, anteromedial, central lateral, central medial, and posterior regions of the thalamus.

White Matter volume

Cui et al. (2020) examined white matter integrity and found that the depression group exhibited more extensive FA reductions compared to the schizophrenia group. Compared to the schizophrenia group the depression group also had significantly reduced FA measurements in the bilateral anterior section of the anterior corona radiata (CR). Results

also showed significantly reduced FA in the anterior and posterior portions of the anterior CR in depression patients compared to patients with schizophrenia. In contrast, the similarities between depression and schizophrenia can be seen in the overlapping regions with decreased FA which were primarily in the posterior section of the anterior CR.

Depression patients who were on medication had significantly lower FA values in the left anterior CR in comparison to depression patients who were not on medication. Finally, the effect of symptom severity and duration of illness on these results did not survive correction for multiple comparisons in voxel-wise and ROI analysis.

Kochunov et al. (2013) showed that a quadratic diagnosis by age interaction was significant in five white matter tracts with the impact of the diagnosis of schizophrenia by age factor indicating that the rate of age-related rate decline was twice higher than in the healthy control group. Also results in the schizophrenia cohort showed that higher age-related decline in FA was present in all twelve tracts analysed. Specifically, the age-of-intercept for average FA was 32.4 years but regionally it was from 27.7 years for the genu of the CC to 49.9 years for the thalamic radiation. The age-of-intercept was negatively correlated with normative age of FA peak suggesting earlier onset of accelerated ageing in FA in schizophrenia in tracts that are normally matured slowly. On the other hand, in the depression group, patients demonstrated higher age-related decline in FA for seven tracts compared to HC of which none were statistically significant (genu of CC, body of CC, internal capsule, external capsule, Cingulum, SLF and inferior frontal occipital). Therefore, this suggests no evidence of tract-specific accelerated ageing in the normal lifespan of depression patients.

Finally, Ota et al. (2013) found that patients with depression exhibited reduced FA in the CC, bilateral anterior and posterior insula, anterolateral and posterior regions of the thalamus, and left rostral and right subcallosal regions of the thalamus.

Discussion

This is the first systematic review to summarise the results from structural brain imaging studies directly comparing patients with psychosis and depression. Our findings suggest that patients show GMV and WMV reductions in a number of similar areas, with some studies suggesting that GMV loss in the insula and thalamus may have some specificity to psychosis. Studies report WMV loss in depression and psychosis, with some potential for specificity in CR and CC FA reductions to depression.

Five studies reported on GMV in depression and psychosis. Overall, they suggest that patients with psychosis and depression show reductions in GMV in the left-sided posterior subgenual prefrontal cortex, the MFG, and the hippocampus. In schizophrenia, the greatest reductions are found in the MFG. In one study, an increase in the posterior subgenual prefrontal cortex volume was found only in the depression group upon follow-up. Some of these studies reported no significant differences in GMV between the two groups: including no significant differences in the female brain, subgenual prefrontal cortex as well as one study reporting no hippocampal GMV volume differences between both groups. Four studies reported on WMV and a transdiagnostic pattern of decreased WMV in both depression and psychosis was identified. This included FA reductions and WMV decline. Two studies found significantly reduced FA specifically in depression. This suggests that both psychosis and depression patients display significantly reduced patterns of WMV, with some studies indicating this is more common in depression patients.

Existing literature has demonstrated that depression and psychosis share common etiologies and symptoms which indicates that both disorders may be presenting transdiagnostic mechanisms (Correll & Schooler, 2020; Upthegrove et al., 2017). The literature for amygdala and hippocampal volume in depression and psychosis is conflicting, similar to our

results. Part of Keller *et al.* (2008) results align with a systematic review which also found no significant differences between the two disorders when examining the hippocampus (O'Connor & Agius, 2015). Keller *et al.* (2008) found volume reductions in the right and left amygdala whereas O'Connor and Agius (2015) found no distinctions. They suggest that reductions in amygdala volume in patients with psychosis contribute to adverse implications such as severe anxiety and fear (Keller *et al.*, 2008). Reduced hippocampal volume has also been closely linked to psychosis in the literature (Haukvik *et al.*, 2018), causing implications such as intensified symptom severity and impaired learning and memory functioning (Arnold *et al.*, 2015). This supports Meisenzahl *et al.*'s (2010) findings where patients with psychosis displayed a greater reduction in hippocampal volume compared to patients with depression.

Research comparing depression and psychosis directly is limited, especially in regards to brain structure, suggesting further evidence is needed to disentangle transdiagnostic signatures and specificity of each of the two disorders. Although comorbidities of depression and psychosis are common (Correll and Schooler, 2020; Kambeitz *et al.*, 2015), the existing literature is sparse when investigating brain structure and its impact on diagnosis, symptom severity, and prognosis. Our results highlight that patients with depression and psychosis exhibit reductions in GMV and WMV with GMV loss being more prominent in psychosis. Coryell *et al.*, 2006, however, found that patients with psychotic depression exhibited more GMV loss than patients with schizophrenia – a result going against the GMV loss hierarchy suggested by our results.

Our findings are in line with previous results showing WMV and FA disparities in both schizophrenia (Kelly *et al.*, 2018) and depression (J. Jiang *et al.*, 2017; Wise *et al.*, 2016). Furthermore, literature has also suggested that GMV reductions have been found in both psychosis and depression, such as (Arnone *et al.*, 2012) systematic review on depression,

which found both WMV and GMV reductions in patients with depression as well as a significant amount of WMV lesions which have been associated with depression and cognitive decline. Moreover, (Steen et al., 2006) meta-analysis documented that patients with first episode psychosis experienced GMV loss, with continuous decline found over time. Subsequent systematic reviews confirmed this, with reduced GMV being found in both patients with schizoaffective disorder and patients with schizophrenia (Madre et al., 2016). A novelty in our findings is that GMV reductions were suggested to be more prevalent in psychosis, with WMV reductions being more prevalent in depression. This finding should however be approached with caution, as some studies did not follow this pattern such as Coryell (2005) who found GMV reductions in both depression and schizophrenia compared to healthy controls and volume loss in the left-sided posterior subgenual prefrontal cortex being significantly lower in the depression group.

The relevance of WMV reductions in the aetiology of depression still remains in question. WMV in patients with depression has been shown to increase after serotonin and norepinephrine reuptake inhibitor (SNRI) or selective serotonin reuptake inhibitor (SSRI) medication compared to recovered patients and healthy controls (Zeng et al., 2012). This implies that SSRIs and SNRIs could possibly reverse WMV reductions, suggesting that WMV may play a role in the aetiology of depression (Zeng et al., 2012).

On the contrary, medication for schizophrenia has shown to cause further impairment in brain structure (Zhang et al., 2018; Zhang et al., 2016a). Studies have demonstrated how antipsychotic drug treatment can further reduce GMV as it can cause neuronal and cell complications (Zhang et al., 2016). However, it can be argued that GMV reductions may be a result of the disorder's severity and need for higher doses of medication (Zhang et al., 2016). Yet, studies show that antipsychotic medication for schizophrenia increases GMV in the

thalamus, which modulates perception, thoughts, and feelings (Yue et al., 2016), resulting in improvements in positive symptoms.

To date, this is the first systematic review which has provided an overview of structural neuroimaging brain comparisons between depression and psychosis. Key strengths of this systematic review include that it reduced reviewer bias as its approach of validly assessing individual studies is objective and reproducible (Bero, 2006; Collins & Fauser, 2005), it used a standardised search method in multiple databases, it was pre-registered, authors were consistently cross-referencing, and efficient quality checks were performed. However, limitations included the lack of coordinate information which prevented the conduction of a meta-analysis, the lack of individual studies directly comparing depression and psychosis, lack of controls for medication exposure, small sample sizes, and heterogeneity of symptoms. Four out of the 7 studies (Coryell et al., 2005; Kochunov et al., 2013; Ota et al, 2013; Cui et al., 2020) scored 4 on the NOS guidelines, indicating low quality. Consequently, future research should commit to a high standard of practice to ensure adequate studies are conducted, in order to provide a collection of satisfactory results and more definitive evidence.

Conclusions

We identified a pattern of GMV reductions and WMV reductions in areas such as the MFG, hippocampus, and left-sided posterior subgenual prefrontal cortex in both depression and psychosis. GMV loss was seen specifically in the insula and thalamus in psychosis. However, the lack of literature directly comparing brain structure in depression and psychosis is further heightened by the poor quality of the studies and future high-quality studies, including those utilising large data sets is needed.

Study	Study Design	Sample size (N)	Age (M, (SD)) and duration of illness (M, (SD))	Patient medication	Setting	Mean symptom score	Type of neuroimaging	Coordinates (MNI or Talairach)
Coryell et al. (2005)	Case-Control	PSDE: 10 SCZ: 10 HC: 10	PSDE Age: 21.9 (4.9) PSDE Duration of illness (years): 4.7 (5.7) SCZ age: 22.2 (4.2) SCZ Duration of illness (years): 5.8 (6.0) HC age: 22.1 (6.0)	N/A	Iowa Longitudinal Study of the Outcome of Early Psychosis	N/A	MRI	N/A
Cui et al. (2020)	Case-Control	MDD: 85 BD: 42 SCZ: 68 HC: 95	MDD Age: 32.4 (8.7) MDD Duration of illness (months): 40.8 (55.8) SCZ Age: 28.5 (4.6) SCZ Duration of illness (months): 43.8 (43.6) HC mean age: 30.2 (6.8)	MDD: 5 AP, 64 AD, 1 MS BD: 24 AP, 15 AD, 24 MS SCZ: 42 AP, 0 AD, 0 MS	Xinxiang Medical University hospital	PANSS-P (SCZ): 22.80 (3.79) PANSS-N (SCZ): 19.97 (5.22) BAI (MDD): 36.69 (12.12) BDI (MDD): 18.97 (7.24)	MRI for microstructural changes in WM	MNI coordinates
Jiang et al. (2019)	Case-Control	SCZ: 20 MDD: 20 HC: 20	SCZ Age: 40.3 (13.8) MDD Age: 41.8 (14.2) HC Age: 41.4 (13.6)	SCZ: 20 ATAP and TAP MDD: 1 TCA, 15 SSRIs	Chengdu Mental Health Centre	PANSS-P (SCZ): 12.9 (5.6) PANSS-N (SCZ): 18.0 (7.0) PANSS-G (SCZ): 27.8 (5.3) HAM-D (MDD): 5.3 (1.3)	VBM analysis for whole brain GMV	MNI coordinates
Keller et al. (2008)	Case-Control	PSDE: 23 NPSDE: 19 HC: 22	PSDE Age: 26.5 (13.2) NPSDE Age: 36.6 (11.9) HC Age: 32.2 (11.5)	PSDE: 16 AP, 15 AD, 4 MS NPDE: 0 AP, 11 AD, 2 MS	Stanford University or online and print advertisement	HAM-D (PSDE): 30.5 (4.7) HAM-D (NPSDE): 23.7 (3.2) BPRS (PSDE): 48.9 (7.8) BPRS (NPSDE): 33.2 (2.8)	MRI	Talairach coordinates
Kochunov et al. (2013)	Case-control	SCZ: 58 HC: 60 MDD: 132 HC: 351	SCZ Age: 37.1 (12.4) HC Age: 37.5 (12.1) MDD Age: 42.3 (12.1) HC Age: 41.5 (13.9)	SCZ: 3 TAP Rest of patients on ATAP or combination of TAP and ATAP 2 medication-free	SCZ and HC: Maryland Psychiatric Research Centre MDD and HC: Genetics of Brain Structure and Function (GOBS) study.	BPRS (SCZ): 15.9 (11.7) BDI (MDD): 11.2 (9.0)	Fractional Anisotropy	N/A
Meisenzahl et al. (2010)	Case-control	SCZ: 89 MDD: 92 HC: 138	SCZ Age: 30.6 (9.1) MDD Age: 44.6 (12.3) HC Age: 33.3 (12.2) SCZ Duration of illness (years): 4.7 (5.9) MDD Duration of illness (years): 5.7 (7.5)	SCZ: 44 TAP, 63 ATAP, 3 medication-free MDD: 21 SSRI, 23 TCA, 42 other AD, 3 medication-free	Department of Psychiatry and Psychotherapy, Ludwig-Maximilians University, Munich, Germany	PANSS-G (SCZ): 75.1 (22.1) HAM-D (MDD): 23.5 (6.7)	MRI	N/A
Ota et al. (2013)	Case-control	SCZ: 43 MDD: 41	SCZ Age: 38.6 (9.6) MDD Age: 41.9 (19.9) SCZ Duration of illness (years): 16.3 (6.1) MDD Duration of illness (years): 8.1 (7.5)	SCZ: CPZ 340.6 (515.2) MDD: IMIE 102.6 (123.0)	N/A	PANSS-P (SCZ): 15.0 (6.0) PANSS-N (SCZ): 15.4 (6.6) PANSS-G (SCZ): 31.0 (10.4) HAM-D (MDD): 12.0 (7.7)	VBM and Fractional Anisotropy	N/A

Table 1: Data extraction table. GMV: grey matter volume; WMV: white matter volume; SCZ: schizophrenia; MDD: major depressive disorder; BP: Bipolar Disorder; HC: healthy controls; ROI: region of interest; FA: Fractional Anisotropy; SSRIs: selective serotonin reuptake inhibitors; F: female; M: male; FA: fractional anisotropy; HAM-D: Hamilton rating scale of depression; PANSS: Positive and negative syndrome scale; SD: standard deviation; AD: antidepressants; BDI: Beck's depression inventory; BAI: Beck's anxiety inventory; IMIE: Imipramine equivalent; MS: Mood stabilizers; TCA: tricyclic antidepressants; TAP: Typical antipsychotics; ATAP

Chapter 4

Heterogeneity and Classification of Recent Onset Psychosis and Depression: a Multimodal Machine Learning Approach

Paris Alexandros Lalousis, MSc^{1,2}; Stephen J. Wood, PhD^{1,3,4}; Lianne Schmaal, PhD^{3,4}; Katharine Chisholm, PhD^{1,5}; Sian Lowri Griffiths, PhD^{1,2}; Renate L.E.P Reniers, PhD^{1,2,6}; Alessandro Bertolino, MD⁷; Stefan Borgwardt, MD⁸; Paolo Brambilla, MD^{9,10}; Joseph Kambeitz, MD¹¹; Rebekka Lencer, MD, PhD¹²; Christos Pantelis, MB BS, MD, MRCPsych, FRANZCP¹³; Stephan Ruhrmann, MD¹⁴; Raimo K. R. Salokangas, MD, PhD, MSc¹⁵; Frauke Schultze-Lutter, PhD¹⁶; Carolina Bonivento, PhD¹⁷; Dominic Dwyer, PhD¹¹; Adele Ferro, PsyD, PhD¹⁰; Theresa Haidl, MD¹⁴; Marlene Rosen, MSc¹⁴; Andre Schmidt, PhD⁸; Eva Meisenzahl, MD¹⁶; Nikolaos Koutsouleris, MD^{11*}; Rachel Upthegrove, MBBS FRCPsych, PhD^{1,2,18*} and the PRONIA Consortium#

* Joint Senior Authors

#See PRONIA consortium author list in the supplement

Author Affiliations:

1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom

3 -Orygen, the National Centre of Excellence in Youth Mental Health; Melbourne, Australia

4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

5 -Department of Psychology, Aston University, United Kingdom

6 -Institute of Clinical Sciences, University of Birmingham, United Kingdom

7 -Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy

8 -Department of Psychiatry, University of Basel, Basel, Switzerland

9 -Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy

10 - Department of Neurosciences and Mental Health, IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

11 -Department of Psychiatry and Psychotherapy, Ludwig Maxmilians University, Munich, Germany

12 -Department of Psychiatry, University of Münster, Münster, Germany

13 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia

14 -Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany

15 -Department of Psychiatry, University of Turku, Turku, Finland

16 -Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

17 -IRCCS “E. Medea” Scientific Institute, San Vito al Tagliamento (Pn), Italy

18 -Early Intervention Service, Birmingham Women's and Children's NHS Foundation trust, United Kingdom

Abstract

Diagnostic heterogeneity within and across psychotic and affective disorders challenges accurate treatment selection, particularly in early stages. Delineation of shared and distinct illness features at the phenotypic and brain levels may inform the development of more precise differential diagnostic tools. We aimed to identify prototypes of depression and psychosis to investigate their heterogeneity, with common, comorbid transdiagnostic symptoms. Analysing clinical/neurocognitive and grey matter volume (GMV) data from the PRONIA database, we generated prototypic models of recent-onset depression (ROD) vs. recent-onset psychosis (ROP) by training support-vector machines to separate patients with ROD from patients with ROP, who were selected for absent comorbid features (pure groups). Then, models were applied to patients with comorbidity, i.e., ROP with depressive symptoms (ROP+D) and ROD participants with sub-threshold psychosis-like features (ROD+P), to measure their positions within the affective-psychotic continuum. All models were independently validated in a replication sample. Comorbid patients were positioned between pure groups, with ROP+D patients being more frequently classified as ROD compared to pure ROP patients (clinical/neurocognitive model: $\chi^2=14.874$; $p<0.001$; GMV model: $\chi^2=4.933$; $p=0.026$). ROD+P patient classification did not differ from ROD (clinical/neurocognitive model: $\chi^2=1.956$; $p=0.162$; GMV model: $\chi^2=0.005$; $p=0.943$). Clinical/neurocognitive and neuroanatomical models demonstrated separability of prototypic depression from psychosis. The shift of comorbid patients towards the depression prototype, observed at the clinical and biological levels, suggests that psychosis with affective comorbidity aligns more strongly to depressive rather than psychotic disease processes. Future studies should assess how these quantitative measures of comorbidity predict outcomes and individual responses to stratified therapeutic interventions.

Introduction:

Treatments for mental illness are currently based on categorical structures built on patterns of syndromes and their course, rather than aetiology (Bzdok & Meyer-Lindenberg, 2018). The biological and clinical overlaps between these syndromes, and significant heterogeneity in outcomes, has become more apparent in recent years (Kambeitz et al., 2015, 2017; Upthegrove et al., 2020). Advancement in both pharmacological and psychotherapeutic interventions has stalled, potentially as a result of continued focus on invalid disease categories (Hofmann et al., 2012; Wong et al., 2010). The need for better treatments is particularly acute in psychosis and depression, which constitute major mental health challenges to the world's population (Charlson et al., 2018; Chong et al., 2016; Conley et al., 2007; Dieleman et al., 2016; Olesen et al., 2012; Sobocki et al., 2006; *WHO | Depression and Other Common Mental Disorders*, 2017). The legacy of a Jaspers based hierarchical approach to symptom structures suggests that positive psychotic symptoms are of primary importance within psychotic spectrum disorders (Bürgy, 2008). Yet the categorical and hierarchical division of psychotic disorders into affective and non-affective has been contested for decades (Quattrone et al., 2019), with clear demonstration of the presence of affective symptoms in psychosis and psychotic symptoms in affective disorders (Conley et al., 2007; Ohayon & Schatzberg, 2002; *WHO | Depression and Other Common Mental Disorders*, 2017).

Heterogeneity is particularly noticeable in early and developing stages of illness, with high prevalence of affective symptoms across disorders (Rickwood et al., 2015). The comorbidity of depression in early psychosis has been largely regarded as secondary to the primary disorder (psychosis), reinforcing a categorical, hierarchical approach (Birchwood et al., 2005; Foulds & Bedford, 1975). There are increasing calls to use empirical evidence to develop alternative aetiologically informed structures (Clark et al., 2017). The use of

multidimensional item response modelling to predict psychosis biotypes has been shown to transcend traditional diagnostic boundaries; with suggestion of an underlying transdiagnostic dimension across psychotic diagnoses (Dwyer et al., 2020; Maj, 2016; Tamminga et al., 2013). However, there are valid reasons why a categorical approach to mental illness has persisted; a significant number of individuals' presentation will 'fit' within distinct categories of mental illness and the course and outcome of their treatment can be predicted from such diagnostic structures (Potuzak et al., 2012). When disorders are fully formed, course as well as symptom structure enable clearer distinction between categories. Imaging studies also show shared areas of interest, including the hippocampus and cerebellum (Meisenzahl et al., 2010), the prefrontal cortex and insula (Busatto, 2013) and both depression and psychosis have been associated with heightened brain activation in regions central to emotional processing (Broome et al., 2015; Busatto, 2013; Kumari et al., 2016) with similarities particularly prominent in the early stages of illness (Broome et al., 2015; Busatto, 2013; Koutsouleris et al., 2015; Kumari et al., 2016).

The distinction between depression and schizophrenia is possible by structural brain data, but also, more challenging in early stages of illness when symptoms and course are more heterogeneous (Koutsouleris et al., 2015). Transdiagnostic processes of mental health disorders are descriptively transdiagnostic (i.e. being present in multiple disorders, without regard to how or why) or mechanistically transdiagnostic (i.e. reflecting neurobiological, physiological, or functional mechanisms) (Harvey et al., 2004, 2011). Both depression and psychosis are associated with transdiagnostic features of working memory, executive functioning, and verbal fluency deficits (Carrà et al., 2019; Moritz et al., 2002). The importance of certain mechanistically transdiagnostic symptoms is potentially hidden in categorical structures, and they remain under-investigated (Krueger & Eaton, 2015).

Complex psychopathology and heterogeneity in developing mental health disorders presents the opportunity of fuller exploration of the significance of potential transdiagnostic symptoms, to provide further insight into aetiopathogenetic pathways of symptoms and through this to advance diagnostic structures (Dwyer et al., 2018; Klöppel et al., 2012). However, novel approaches and powerful statistical tools such as machine learning techniques could help provide this deeper understanding by detecting complex patterns of data across diagnostic structures, and the delineation of shared and distinct features of these illnesses at the phenotypic and brain levels. This may inform the development of more precise differential diagnostic tools and improve the development of new treatments (Dwyer et al., 2018).

This study aimed to identify prototypes of pure depression and pure psychosis in order to investigate the heterogeneity of depression and psychosis with common, comorbid transdiagnostic symptoms. We hypothesized that developed models would correctly classify diagnostic groups without comorbid symptoms, in keeping with evidence of the utility in categorical diagnostic structure, and that grey matter volume (GMV) would add classification accuracy. We further hypothesised that a reduction of classification accuracy would be seen in groups with comorbid symptoms. Exploration and the delineation of shared and distinct features at both the phenotypic and biological levels may potentially inform future development of more precise treatments.

Materials and Methods

Study design

Data were taken from the discovery and replication samples of the PRONIA study, an EU-FP7 funded seven centre study aiming to optimize candidate biomarkers for the prediction and staging of mental health disorders. Details of the PRONIA study sites, recruitment

protocol and quality control procedures are described in a previous publication(Koutsouleris et al., 2018b) and in the supplementary methods (1.1, 1.2, 1.3, tables S2, S3).

Inclusion and Exclusion Criteria

The general inclusion criteria for the study were: (1) age between 15 and 40 years, (2) sufficient language skills for participation, (3) capacity to provide informed consent/assent. General exclusion criteria were: (1) an IQ below 70, (2) current or past head trauma with loss of consciousness (> 5 minutes), (3) current or past known neurological or somatic disorders potentially affecting structure or functioning of the brain, (4) current or past alcohol dependence, (5) polysubstance dependence within the past six months, and (6) any medical indication against MRI. ROP and ROD inclusion criteria can be found in the supplement (1.1).

Group identification

Pure ROP: any ROP patient who had a Beck Depression Inventory-II (BDI-II) score of 13 or lower, which is indicative of absent or minimal depressive symptoms (A. Beck et al., 1996; A. T. Beck et al., 1996). Pure ROD: any ROD patient who had a Positive and Negative Symptom Scale (PANSS) (Kay et al., 1987) positive subscale score of no more than 7 and no Structured Interview of Psychosis-risk Syndromes positive (SIPS-P) severity score of 3 or more on any item.

ROP with depressive symptoms (ROP+D): any ROP participant with a BDI-II score of 14 or more. ROD with psychotic symptoms (ROD+P): any ROD participant with a SIPS positive item score of 2 or more and a Schizophrenia Proneness Instrument-Adult version (SPI-A) Cognitive Disturbances (COGDIS) item score of 3 or more.

Twenty-four ROP and 31 ROD patients from the discovery sample and 21 ROP and 53 ROD patients were not included in the analysis due to not meeting group identification criteria or not having neuroimaging data.

MRI imaging data acquisition, quality control, and preprocessing

Participants underwent a multi-modal MRI protocol. A minimal harmonization protocol, which the MR sequences across the different scanners had to comply with is described in the supplementary methods (1.3). In the current study, T1-weighted structural MRI (sMRI) images of the participants were analyzed. The sMRI images of six healthy travelling volunteers who were scanned at all sites with same parameters were also analysed as part of a calibration study. The images were processed using the open source CAT12 toolbox (version r1155; <http://dbm.neuro.uni.jena.de/cat12/>) (see supplementary methods 1.4). Employing generalization theory (Mushquash & O'Connor, 2006; Robert L. Brennan, 2011), a between-site voxel reliability map (G coefficient map) was computed from the analysis of the GMV maps that were derived from the calibration study. During our neuroimaging based machine learning analyses the G coefficient maps were used for reliability-based voxel masking.

MRI harmonization and data acquisition

When setting up the PRONIA study, we decided to generate a MRI database that would represent the MR scanner sequence heterogeneity encountered in clinical real-world. The aim of this strategy was to strengthen the generalizability and clinical applicability of the predictive models developed by our machine learning analyses. Thus, we agreed on a minimal harmonization protocol that required the PRONIA sites to only (1) acquire isotropic or nearly isotropic voxel sizes of preferably 1 mm resolution, (2) set the Field Of View (FOV) parameters accordingly to guarantee the full 3D coverage of the brain including all parts of the cerebellum, and (3) define the relaxation time (TR) and echo time (TE) as well as

other imaging parameters in a way that would maximize the contrast between cortical ribbon and the white matter and enhance the signal-to-noise ratio in the images. At every site all the images were visually inspected, automatically defaced and anonymized using an in-house Freesurfer-based script before the data was centralized.

MRI processing pipeline

The manual of the CAT12 toolbox (<http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>) details the processing steps applied to the structural images. These steps consist of:

- (1) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering(Manjón et al., 2008).
- (2) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local variations of intensity distributions as slowly varying spatial functions and thus achieves a homogeneous segmentation across cortical and subcortical structures(Rajapakse et al., 1997).
- (3) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior information of adjacent voxels into the segmentation estimation generated by AMAP(Rajapakse et al., 1997).
- (4) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM) inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP segmentation.
- (5) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the AMAP-generated tissue segments.

- (6) A high-dimensional DARTEL registration of the image to a MNI-template generated from the MRI data of 555 healthy controls in the IXI database (<http://www.braindevelopment.org>).
- (7) The GM maps were then multiplied with the Jacobian determinants that were obtained during registration in order to produce GM volume maps.
- (8) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

Classification Models

Using the pure ROP and ROD groups, a Support Vector Machine (SVM) classification model was built using individual item scores from broad clinical and neurocognitive tests that assess features commonly occurring in psychosis and depression including anhedonia, social functioning and cognition deficits (see supplementary methods 1.8). In total, 151 features were included in the model. The trained pure classification model was then applied to the comorbid ROP+D and ROD+P groups to determine the classification accuracy to their primary diagnosis.

A second model using GMV whole-brain voxel-wise data as features was developed in the pure ROD and ROP groups and then applied to the ROP+D and ROD+ P groups. The developed clinical/cognitive and GMV models were combined by using decision values from the pure clinical and pure GMV models (in order to build a model that learns from the meta-data) in a stacking-based data fusion framework (Polikar, 2006; Wolpert, 1992). Finally, all the models were applied to an independent replication sample.

Support Vector Machine Learning Analysis:

The machine learning analysis of pre-processed data (see supplementary methods 1.5) was performed using NeuroMiner (version 1.0; <https://github.com/neurominer-git>). A repeated nested pooled cross-validation (CV) was used with 10 outer CV2 permutations, 10 outer CV2 folds, 10 inner CV1 permutations, and 10 inner CV1 folds.

Imbalanced learning was corrected for, by increasing the C value in the minority class by multi-plying it by the inverse ratio of the training class sizes. A linear kernel was used with eleven C values (0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16) in order to optimize the choice of C value and create an ensemble of predictive models to be applied to the CV₂ data to produce a single average robust prediction.

Balanced accuracy (BAC) regularized by SVM model complexity was used as a criterion for the hyperparameter optimization (see supplementary methods 1.6).

Finally, a stacking-based data fusion framework (Karrer et al., 2019; R. Polikar, 2006; Sabbagh et al., 2020; Wolpert, 1992) was used to examine whether the combination of the clinical/neurocognitive and the neuroimaging-based models would provide a superior classification accuracy (see supplementary methods 1.7).

Independent Validation

All our models were validated in our independent replication sample (N=262) (see supplementary methods table S6) which was collected at a different timescale. The same group identification criteria were applied.

Supplementary analyses

A number of supplementary exploratory analyses, including correlation analyses between decision scores from our models, association and comparison analyses between correctly and mis-classified patients, GMV comparison between groups, decision score group comparisons, and regression analyses with 9 month functional outcomes were conducted and can be found in the supplement (Section 2).

Results:

Demographic Information

Data from 154 participants with ROP and 146 patients with ROD were included in the analysis as our Discovery sample. Thirty-eight ROP patients were included in the pure ROP group and 90 ROD patients in the pure ROD group. The mean age of the pure ROP group was 26.5 [SD 6.8]) and the mean age of the pure ROD group was 26.5 [SD 6.6]). There were 25 male and 13 female patients in the pure ROP group and 45 male and 45 female patients in the pure ROD group.

Ninety-two ROP subjects were included in the ROP+D group and 25 ROD subjects in the ROD+P group. The mean age of the ROP+D group was 26.5 [SD 6.8] and the mean age of the ROD+P group was 23.8 [SD 3.9]. There were 57 males and 35 females in the ROP+D group and 12 males and 13 females in the ROD+P group. A summary of demographic information is provided in table 1.

The independent validation sample consisted of 161 patients with ROP and 131 patients with ROD. Fifty ROP patients were included in the pure ROP group and 53 ROD patients in the pure ROD group. The ROP+D and ROD+P groups consisted of 90 and 25 patients respectively. A full description is provided in the supplement (Table S6).

Machine Learning Analyses

Internal Validation of the Pure Group Differential Classifiers

Clinical and Neurocognitive Data

A repeated nested pooled cross validation model with classifiers of clinical and cognitive variables predicted pure diagnostic groups with a balanced accuracy (BAC) of 79.3%; 95% CI [77.2, 82.3] and an area under the curve (AUC) of 0.86 (Table 2 and Figure 1a).

Assignment to the ROP category by the clinical classifier was driven by reduced scores in the RSA and elevated scores in the WSS, RSA, and SPIA. ROD group classification was informed by increased scores in the SPIA, WSS, together with reduced scores in the DSST. The contribution of the features was calculated by feature weights and by cross-validation ratio (Figure 2).

GMV Data

The repeated nested pooled cross-validation model using sMRI to predict diagnostic group in pure ROP and ROD produced a BAC of 62.5%; 95% CI [58.8, 64.0] and an AUC of 0.70 (Table 2 and Figure 1b). ROP patients showed pronounced reductions in the thalamus and the cerebellum, whereas depressed patients showed orbitofrontal, limbic and paralimbic volume reductions (Figure 3).

Stacking

Combining the outputs of the clinical predictors and sMRI using stacked generalization predicted diagnostic group with a BAC of 79.5%; 95% CI [77., 81.9] and an area under the curve (AUC) of 0.87 (Table 2 and Figure 1c).

Separability of Comorbid Groups

Clinical/Neurocognitive Data: The trained pure classification system comprising the collection of 11 clinical/neurocognitive models generated by the repeated nested cross-validation scheme on pure groups was then applied to the comorbid groups (ROP+D and ROD+P) to produce decision scores measuring ROP vs. ROD likeness. This model had a BAC of 62.5% and an AUC of 0.66. Misclassifications showed a directionality toward the ROD group, with 63% of ROP+D patients being classified as ROD; $Z=1.276$, $p=.0385$) (see Table 2 and Figure 1a). ROP+D patients were more frequently classified as ROD compared to pure ROP patients ($\chi^2=14.874$; $p<0.001$). In contrast, the assignment precision of ROD+P and ROD patients did not differ ($\chi^2=1.956$; $p=0.162$).

GMV Data

The trained pure classification system (comprising of 11 GMV models generated by the repeated nested cross-validation scheme on pure groups) was then applied to the comorbid groups (ROP+D and ROD+P) to produce decision scores measuring ROP vs. ROD likeness. This produced a BAC of 47.8% and AUC of 0.43. Misclassifications showed a directionality toward the ROD group, with 80.4% of ROP+D patients being classified as ROD; $Z=.713$, $p=.344$) (see Table 2 and Figure 1b). Similarly to the clinical/neurocognitive model ROP+D patients were more frequently classified as ROD compared to pure ROP patients ($\chi^2=4.933$; $p=0.026$). In contrast, the assignment precision of ROD+P and ROD patients did not differ ($\chi^2=0.005$; $p=0.943$).

Stacking

When applied to the comorbid groups, the combined model predicted diagnostic group with a BAC of 58.5% and an area under the curve (AUC) of 0.66 (see Table 1 and Figure 1c).

Independent Validation

Application of our models to the independent validation sample replicated findings very well (pure clinical and neurocognitive model BAC 76.2; pure imaging model BAC 49.9; pure stacking model BAC 78.2). Full results from the independent validation analysis can be found in Table 1.

Supplementary Analyses

See supplement section 2 for additional exploratory analyses results.

Discussion:

Using repeated nested cross-validation techniques we built classification models based on transdiagnostic clinical and neurocognitive features and GMV data, together with a combined model integrating all data modalities, to classify prototype diagnostic groups of ROD and ROP participants without comorbidity. Eighty-seven per cent of patients with pure ROP and ROD were accurately ascribed to their diagnostic group. Applying this model to groups with comorbidity, 88% of patients with ROD and psychotic features were ascribed to their primary diagnostic group (depression) whereas only 37% of patients with ROP and depressive features were ascribed to their primary diagnostic group (psychosis). The shift of comorbid psychosis patients towards the depression prototype was observed both at the clinical and biological levels. This suggests that when comorbid with affective symptoms, psychoses align more strongly to the disease processes of depressive than psychotic disorders. Using GMV measures only for classification, comorbid groups ROP patients with depressive symptoms largely resembled pure ROD. Results were generalisable to our independent validation sample.

Our findings suggest that clinical and neurocognitive transdiagnostic symptoms may have differential weight within psychosis and depression presentation with and without comorbidity. In pure groups, these symptoms could be seen as under the hierarchical umbrella of psychosis or depression, and may accurately reflect underlying pathology in these groups. However, in the face of comorbidity, transdiagnostic symptoms lean more to the depression domain. Given the prevalence and importance of depressive comorbidity in early psychosis this may support a model where depression could be more intrinsically important than is currently considered in early phases of illness (Upthegrove et al., 2017, 2020). When participants with ROP exhibited even mild depressive symptoms, their GMV classification was more likely to lie within the depression group than psychosis suggesting a potential depressive biological phenotype that exists in the psychosis spectrum. The implications of these findings suggest that understanding heterogeneity of brain structures may need to include specific focus on symptoms that may often be masked by more acute (e.g. positive) symptoms and simple solutions that some symptoms may be transdiagnostic, potentially belie the complexity of individual aetiology and psychopathology. Furthermore, these findings indicate a need to rethink current diagnostic classification to better reflect the biological reality and eventually develop better treatment options.

A classical diagnostic hierarchy in the structure of personal illness, reflected in current nosological classification systems, posits that mental disorders of the primary diagnosis have more weight and primacy over symptoms from lower classes, which are seen as secondary or comorbid (Birchwood et al., 2005; Foulks & Bedford, 1975). We found that comorbid symptoms affect diagnostic structures in different ways. Sub-clinical psychotic-like symptoms appeared to not alter the signature of ROD patients whilst depressive symptoms had a profound effect on ROP patients' classification accuracy.

Our findings add evidence to the debate around the validity of the system on which the DSM is built upon (Stein et al., 2010), suggesting that a descriptivist position in the diagnosis of mental disorder is not sufficient and that a novel multivariate approach of mental disorder is more appropriate. Comorbidity in psychiatric disorders presents a clinical as well as nosological challenge. There may be significant interplay between sets or clusters of symptoms over the development of disorder from prodrome via onset to potential chronicity. The frequency of depression within ROP may be a primary driver, rather than being a secondary symptom. If identified correctly, novel symptoms may be new targets that if treated effectively, ameliorate the other, e.g. positive symptoms.

Recent onset disorders may constitute groups of phenotypically highly individual symptoms with underlying aetiopathology, and it may be that personalized treatments could be tailored accordingly. Our SVM classification model found that different types of anhedonia (social and physical) were important in the classification of both psychosis and depression.

Anhedonia has been suggested as a possible biomarker for depression (Der-Avakian & Markou, 2012) and has been found to be associated with decreased activation in ventral basal ganglia areas, the dorsal anterior cingulate, middle frontal gyrus, and medial frontal gyrus both in schizophrenia and depression (Zhang et al., 2016). Our GMV model revealed that orbitofrontal areas were higher weighted in the classification to the ROD group.

The relationship between psychotic and affective symptoms has been central to the dilemma of psychiatric classification. Substantial clinical and genomic evidence shows that schizophrenia and affective disorders may be distributed across a dimensional spectrum (Smeland et al., 2019). However, the dimensional spectra model does not allow for either the clinical reality of a complex and changing symptom profile, nor the investigation of clinical features commonly seen across all disorders; some of which may be of primary importance. Concerning the neurobiology of schizophrenia and depression, the majority of previous

studies are based in subjects with depression or psychosis, but only rarely in both (Koutsouleris et al., 2008, 2015) and not previously in highly mixed recent onset comorbid disorders. Our results suggest that while GMV showed some distinction in prototype (pure) groups, when presented with complex comorbid groups, which may be the majority in clinical practice, there was a significant lack of any point of rarity between disorders. This builds on previous work suggesting distinction is more challenging, but also that neuroimaging based data do not support categorical classification in recent onset disorders (Koutsouleris et al., 2015). In recent onset disorders, early cognitive processes related to depression can both drive other more severe symptoms and/or be seen in isolation: for example, anhedonia could be an early indication of negative symptom clusters or a core feature of co-morbid depression (Barkus & Badcock, 2019).

Strengths and Limitations

The strengths of the present analysis include sufficiently large data, robust collection of clinical and imaging data from both depression and psychosis groups, independent validation analysis together with a novel approach to a challenging and essential clinically relevant research question, which speaks to the validity of diagnoses as the cornerstone of psychiatric practice. Our results however should be interpreted with certain caution due to limitations with the study. Regarding our definition of the ROD+P group we used a SIPS-P item score of 2 or more which is not a marker of formal psychotic symptoms, and thus would only measure low levels of psychotic-like symptoms. However to supplement this we used a SPI-A COGDIS item score of 3 or more. We did not include core symptomatology measures such as the PANSS and the BDI in our features due to the fact that primary groups are defined with these measures, and therefore including them would risk a circular analysis. Finally, there were more subjects identified in the pure ROD and comorbid ROP groups; ideally groups of equal size would have been used. Nevertheless, we addressed this imbalance in our analysis,

by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes.

Conclusions

Findings from this large, multi-modal, replicated machine learning classification study in recent onset disorders suggest that whilst there may be a small subset of prototypically pure individuals with clear categorical disorder, the majority of patients share a number of transdiagnostic features, primarily from the depression domain. Brain structure of psychosis patients with co-morbid depressive symptoms largely resembles that of depression. The increasing interest in heterogeneity of early disorders and transdiagnostic symptoms as novel treatment targets needs to be fully informed of potential depression related co-morbidity. Our analysis in recent onset groups also highlight that both categorical and transdiagnostic approaches may ultimately fail at an individual patient level, as neither recognise the possibility of pluripotent pathways that are both stage and context dependent. The implications, particularly for early intervention and prevention in mental health disorders is that ultimately a personalised medicine approach, encompassing the full potential of comorbidity, may be necessary to improve outcomes. Future studies should investigate the utility of targeting such transdiagnostic depression features to elucidate their prognostic value, and develop new treatments.

	ROP Group	ROD Group	t/χ^2	P Value	Pure ROP Group	Pure ROD Group	t/χ^2	P Value	ROP+D Group	ROD+P Group	t/χ^2	P Value
Sample Sizes, No.	154	146			38	90			92	25		
Participants per site, No. (%)												
Basel	23 (7.7)	17 (5.7)			3 (2.3)	8 (6.9)			14 (12)	6 (5.1)		
Birmingham	10 (4.7)	10 (3.3)			4 (3.1)	0 (0)			10 (8.5)	2 (1.7)		
Cologne	27 (9)	27 (9)			4 (3.1)	19 (14.8)			22 (18.8)	3 (2.6)		
Milan	13 (4.3)	7 (2.3)	$\chi^2 = 8.9$	.257	8 (6.3)	5 (3.9)	$\chi^2 = 21.0$	.002	1 (0.9)	0 (0)	$\chi^2 = 6.1$	.517
Munich	46 (15.3)	47 (15.7)			10 (7.8)	35 (27.3)			31 (26.5)	7 (6)		
Turku	22 (7.3)	13 (4.3)			11 (8.6)	7 (5.5)			7 (6)	3 (2.6)		
Udine	12 (4)	21 (7)			2 (1.6)	12 (9.4)			3 (2.6)	3 (2.6)		
Age, Mean (SD)	24.7 (5.4)	25.5 (6.1)		.229	26.5 (6.8)	26.5 (6.6)		.959	24.6 (4.8)	23.8 (3.9)		.449
Sex (Male/Female)	94/60	67/79	$\chi^2 = 6.9$	.009	25/13	45/45	$\chi^2 = 2.6$	.101	57/35	12/13	$\chi^2 = .81$	.368
Education, mean (SD)	13.9 (2.4)	15.1 (7.5)	$t = 1.85$	.064	13.9 (2.2)	15.8 (9.3)	$t = 1.22$	.224	14.0 (2.6)	13.5 (2.3)	$t = -.838$	.404
Educational years repeated, mean (SD)	1.8 (2.5)	2.3 (2.7)	$t = 1.47$	.141	.83 (.87)	1.1 (1.8)	$t = 1.00$	.319	3.1 (4.8)	5.0 (5.4)	$t = 1.70$	.091
Having partnership most of the time in the year before study inclusion, No. (%)	72 (24.7)	85 (29.2)	$\chi^2 = 4.1$	.036	17 (13.4)	51 (40.2)	$\chi^2 = 1.6$	.194	48 (41.4)	18 (15.5)	$\chi^2 = 2.9$	.085
Population density in living area, mean (SD), habitants/km2	3717.8 (2532.3)	3529.8 (2377.1)	$t = -.544$	.587	4498.3 (2724.6)	3022.1 (2274.3)	$t = -2.640$	.010	3447.9 (2338.5)	5152.1 (2361.8)	$t = 2.547$	.013

Table 1. Sample Sociodemographics. Sample Sizes, Participants per Study Site, Age, Sex, Education, Partnership Status, Population Density.

	True Positive, No.	True Negative, No.	False Positive, No.	False Negative, No.	Balanced Accuracy, %	AUC	Model P Value
Clinical/Neurocognitive Pure Model ROP-ROD	29	74	16	9	79.3	0.86	<0.001
Applied Clinical/Neurocognitive Model ROP+D-ROD+P	36	19	6	56	57.6	0.71	NA
Applied Clinical and Neurocognitive Model Validation Pure ROP- ROD	34	39	14	16	70.8	0.78	NA
Applied Clinical and Neurocognitive Mode Validation ROP+D- ROD+P	27	20	5	63	55	0.56	NA
GMV Pure Model ROP-ROD	15	77	13	23	62.5	0.7	<0.001
Applied GMV Model ROP+D-ROD+P	19	20	5	73	50.3	0.47	NA
Applied GMV Model Validation Pure ROP- ROD	17	38	15	33	52.8	0.59	NA
Applied GMV Model Validation ROP+D- ROD+P	31	17	8	59	51.2	0.6	NA
Combined Model ROP-ROD	30	72	18	8	79.5	0.87	NA
Applied Combined Model ROP+D-ROD+P	49	19	6	43	64.6	0.71	NA
Applied Combined Model Validation Pure ROP-ROD	36	38	15	14	71.8	0.78	NA
Applied Combined Model Validation ROP+D-ROD+P	31	18	7	59	53.2	0.56	NA

Table 2. Classification Performance of the Clinical/Neurocognitive, GMV, and Combined Models and Validation Performance.

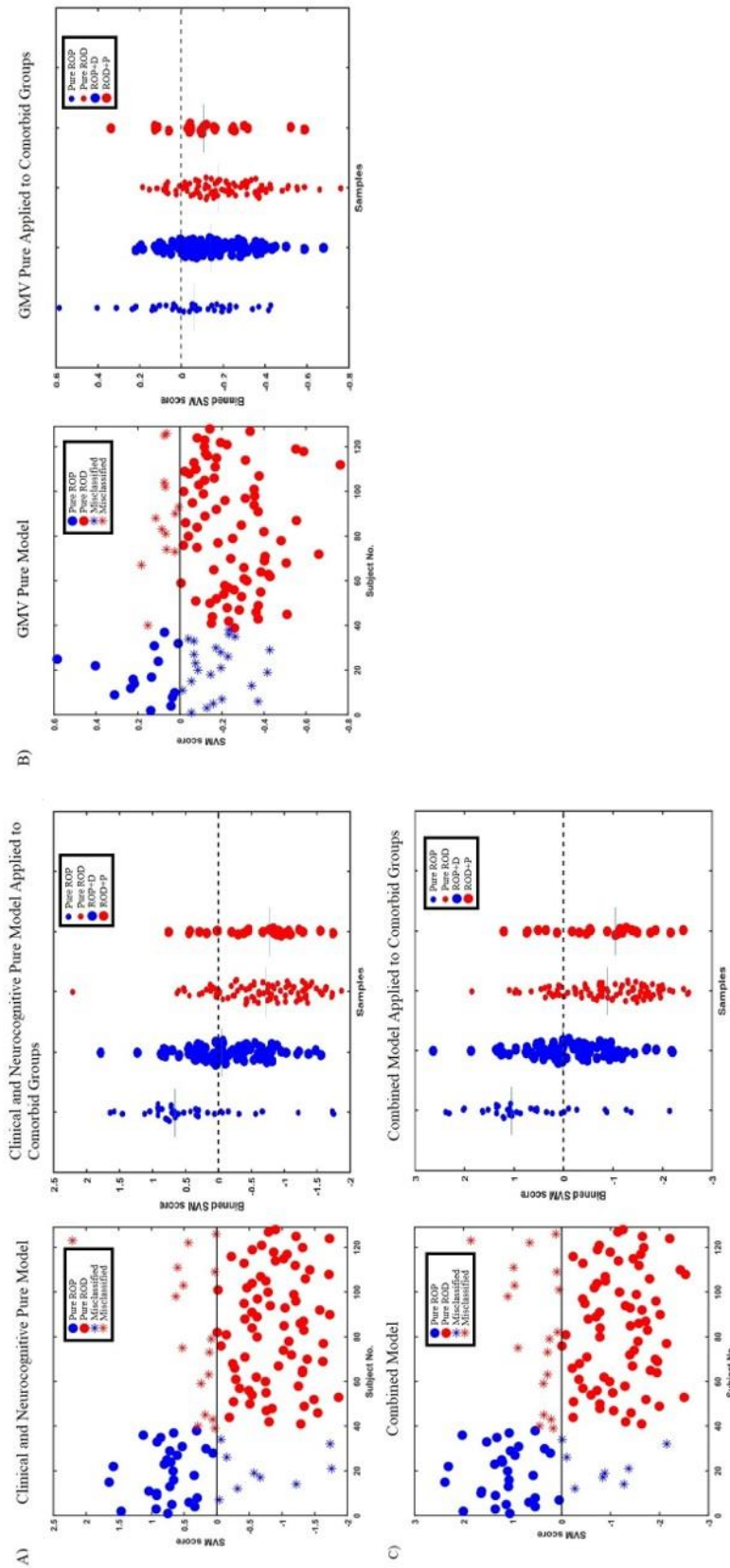
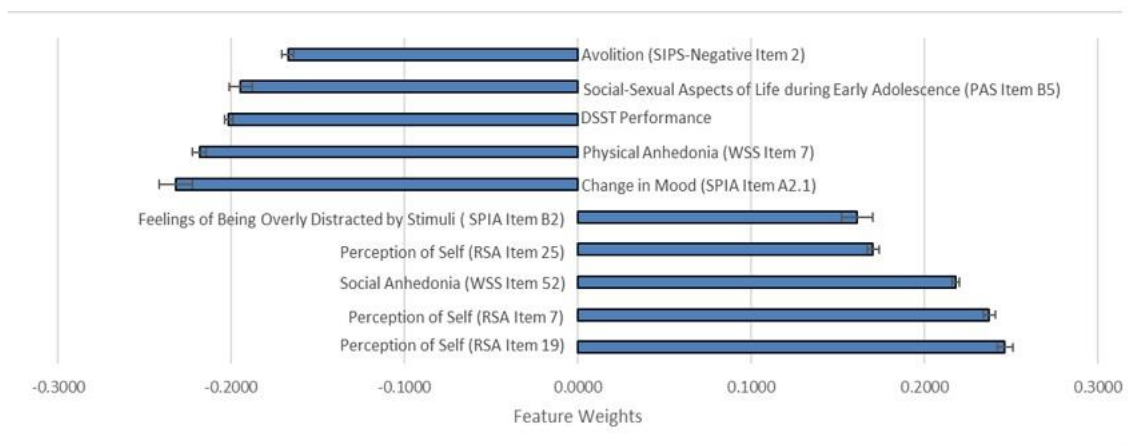
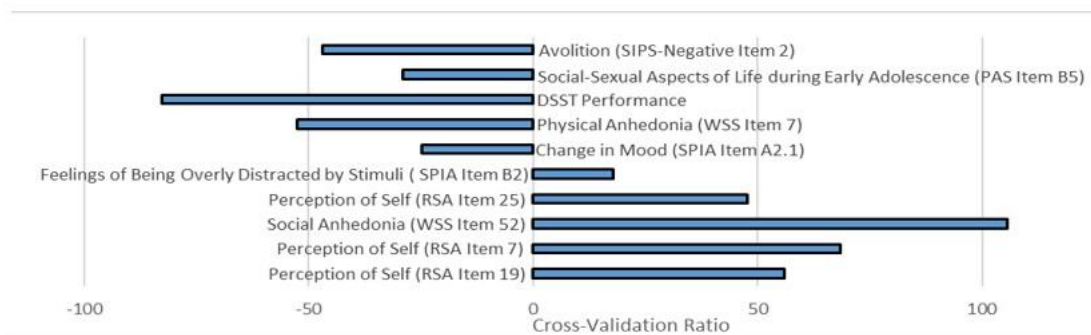


Figure 1. Classification Performance of the Pure and Applied Clinical, GMV, and Combined Models. A) Pure Clinical and Neurocognitive Classification Balanced Accuracy 79.3%, Sensitivity 76.3%, Specificity 82.2%, AUC 0.86. Applied Clinical Classification Balanced Accuracy 57.6%, Sensitivity 39.1%, Specificity 76%, AUC 0.71. B) Classification Performance of the Pure GMV Model and the Applied GMV Model. Pure GMV Classification Balanced Accuracy 62.5%, Sensitivity 39.5%, Specificity 85.6%, AUC 0.70. Applied GMV Classification Balanced Accuracy 50.3%, Sensitivity 20.7%, Specificity 80%, AUC 0.47. C) Classification Performance of the Combined Model and the Applied Combined Model. Stacked Classification Balanced Accuracy 79.5%, Sensitivity 78.9%, Specificity 80%, AUC 0.87. Applied Stacked Classification Balanced Accuracy 64.6%, Sensitivity 53.3%, Specificity 76%, AUC 0.71.



A. Feature Weights. Derived from 1000 random permutations of the outcome labels.



B. Cross-Validation Ratio. Sum of the median weights across all CV1 folds divided by the standard error
Figure 2. Feature Weights and Cross-Validations Ratios of the most Significant Features.

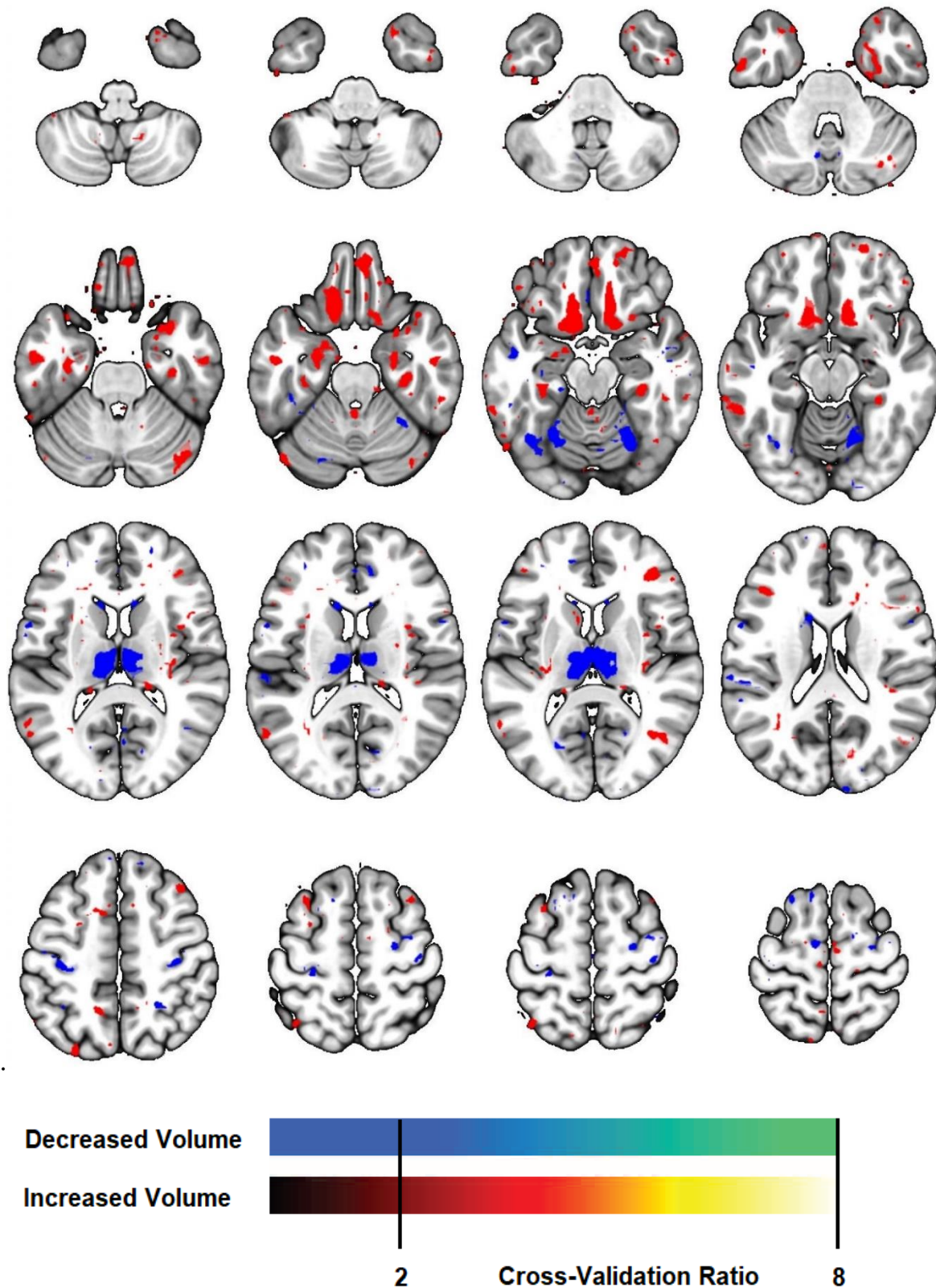


Figure 3. Significant Regions in the Imaging Classification Model. ROP GMV Reductions in the Thalamus and the Cerebellum, ROD GMV Reductions in Orbitofrontal, Limbic, and Paralimbic Regions

Chapter 5

Neurobiologically Based Stratification of Recent Onset Depression and Psychosis:

Identification of Two Distinct Transdiagnostic Phenotypes

Paris Alexandros Lalousis, MSc^{1,2}; Lianne Schmaal, PhD^{3,4}; Stephen J. Wood, PhD^{1,3,4}; Renate L.E.P Reniers, PhD^{1,2,5}; Nicholas M. Barnes, PhD⁵; Katharine Chisholm, PhD^{1,6}; Sian Lowri Griffiths, PhD^{1,2}; Alexandra Stainton, PhD^{3,4}; Junhao Wen, PhD⁷; Gyujoon Hwang, PhD⁷; Christos Davatzikos, PhD⁷; Christina Andreou, MD⁸; Carolina Bonivento, PhD^{9,10}; Udo Dannlowski, MD¹¹; Adele Ferro, PhD^{9,10}; Theresa Liechtenstein, MD¹²; Anita Riecher-Rössler, MD⁸; Georg Romer, MD⁹; Marlene Rosen, PhD¹²; Alessandro Bertolino, MD¹³; Stefan Borgwardt, MD^{8,14}; Paolo Brambilla, MD^{9,10}; Joseph Kambeitz, MD¹⁵; Rebekka Lencer, MD, PhD^{9,12}; Christos Pantelis, MB BS, MD, MRCPsych, FRANZCP¹⁶; Stephan Ruhrmann, MD⁹; Raimo K. R. Salokangas, MD, MSc, PhD, PsD¹⁷; Frauke Schultze-Lutter, PhD^{18,19,20}; André Schmidt, PhD⁹; Eva Meisenzahl, MD¹⁸; Nikolaos Koutsouleris, MD¹⁵; Dominic Dwyer, PhD^{15*}; Rachel Upthegrove, MBBS FRCPsych, PhD^{1,2,21*}, for the PRONIA Consortium#

* Joint Senior Authors

#See PRONIA consortium author list

Author Affiliations:

1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom

3 -Orygen, Parkville, Australia

- 4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia
- 5 -Institute of Clinical Sciences, University of Birmingham, United Kingdom
- 6 -Department of Psychology, Aston University, United Kingdom
- 7 -Perelman School of Medicine University of Pennsylvania, United States of America
- 8 -Department of Psychiatry (UPK), University of Basel, Basel, Switzerland
- 9 -Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy
- 10 - Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy
- 11 –Institute for Translational Psychiatry, University of Münster, Münster, Germany
- 12 -Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany
- 13 -Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy
- 14 -Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany
- 15 -Department of Psychiatry and Psychotherapy, Ludwig Maxmilians University, Munich, Germany
- 16 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia
- 17 -Department of Psychiatry, University of Turku, Turku, Finland
- 18 -Department of Psychiatry and Psychotherapy, University of Düsseldorf, Düsseldorf, Germany
- 19 -Department of Psychology and Mental Health, Faculty of Psychology, Airlangga University, Surabaya, Indonesia
- 20 - University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland

21 -Birmingham Early Interventions Service, Birmingham Women's and Children's NHS
Foundation Trust

Abstract

Background: Identifying neurobiologically based transdiagnostic categories of depression and psychosis may elucidate heterogeneity, and provide better candidates for predictive modelling. We aimed to identify clusters across patients with recent onset depression (ROD) and recent onset psychosis (ROP) based on structural neuroimaging data. We hypothesized that these transdiagnostic clusters would identify patients with poor outcome and allow more accurate prediction of symptomatic remission than traditional diagnostic structures.

Methods: HYDRA (Heterogeneity through Discriminant Analysis) was trained on whole brain volumetric measures from 577 participants from the discovery sample of the multi-site PRONIA study to identify neurobiologically driven clusters which were then externally validated in the PRONIA replication sample (n=404) and three datasets of chronic samples (COBRE, n=146; MCIC, n=202; MUC, n=470).

Results: The optimal clustering solution was two transdiagnostic clusters (Cluster 1, n=153, 67 ROP, 86 ROD and Cluster 2, n=149, 88 ROP, 61 ROD; ARI=.618). The two clusters contained both ROP and ROD. One cluster had widespread GMV deficits, more positive, negative, and functional deficits (impaired cluster) and one cluster revealed a more preserved neuroanatomical signature and more ‘core’ depressive symptomatology (preserved cluster). The clustering solution was internally and externally validated and assessed for clinical utility in predicting 9-month symptomatic remission -outperforming traditional diagnostic structures.

Conclusions: We identified two transdiagnostic neuroanatomically informed clusters which are clinically and biologically distinct, challenging current diagnostic boundaries in recent onset mental health disorders. These results may aid understanding of aetiology of poor outcome patients transdiagnostically and improve development of stratified treatments.

Introduction

The current classification of mental disorders is based on a phenomenological approach that uses signs and symptoms to assign a diagnosis. Whilst some diagnoses have high reliability, their usefulness and aetiopathogenetic basis is questionable (Cuthbert, 2014; First et al., 2018; Fusar-Poli, Solmi, et al., 2019). For example, there is considerable commonality of symptoms and neurobiological domains across mental disorders and co-morbidity frequently occurs; with a prevalence of depression in over 40% of people with schizophrenia (Conley et al., 2007; Upthegrove et al., 2017) and psychotic symptoms occurring in around 20% of people with depression (Johnson et al., 1991; Ohayon & Schatzberg, 2002).

In terms of brain structure, grey matter volume (GMV) reduction is found in both depression and psychosis, across similar areas such as the anterior insula and the dorsal anterior cingulate cortex (Goodkind et al., 2015). This GMV loss has been shown to predate medication exposure, poor functional outcome, neurocognitive deficits, and in the case of clinical high risk for psychosis, transition to frank illness (Meisenzahl et al., 2010; Pantelis et al., 2009; Reniers et al., 2017; Upthegrove et al., 2017). Symptoms common to depression and schizophrenia, such as social withdrawal, blunted affect, and alogia, are associated with GMV reduction in the cerebellum, while anhedonia and avolition are negatively correlated with left anterior limb of internal capsule white-matter volume (WMV) and positively correlated with left superior longitudinal fasciculus WMV (Chuang et al., 2014).

GMV loss in psychosis and depression may be related to immune dysfunction. Elevated pro-inflammatory cytokines, potentially resulting from genomic predisposition or response to environmental factors, may lead to activation of astrocytic dysfunction and/or microglia activation, resulting in dendritic pruning and synaptic changes (Corsi-Zuelli & Deakin, 2021; Laskaris et al., 2021; Sekar et al., 2016). Indeed, immune dysfunction is implicated in the

aetiology of both schizophrenia and depression with cytokines such as IL-6 and CRP detected at elevated levels (Khandaker et al., 2014, 2015; Noto et al., 2015; Osimo et al., 2020; Upthegrove et al., 2014), and causality suggested in mendelian randomisation studies of both disorders (Khandaker et al., 2015; Perry et al., 2021).

Currently, diagnoses are not based on underlying brain structure or distinct biological aetiology. Patients whose symptoms are potentially caused by different biological processes may be given the same diagnosis and patients whose symptoms are potentially caused by same biological processes may be provided with a different diagnosis, a practice which may have detrimental effects on outcome prediction development (Linden, 2012; Stephan, Bach, et al., 2016; Stephan, Binder, et al., 2016). Recent research has highlighted this mismatch between diagnostic labels and the clinical and neuroanatomical picture in depression and psychosis (Lalousis et al., 2021) and heterogeneity may be particularly pronounced in early stages of developing mental health disorders (Baca-Garcia et al., 2007; Keshavan & Brady, 2011; Koutsouleris et al., 2015; Pope et al., 2013; Salvatore et al., 2013). The lack of biological validity of diagnostic groups is thought to be one of the major reasons for poor biomedical translation in psychiatry (Brückl et al., 2020; Kapur et al., 2012; Kendler, 2013).

Only 20% of people with psychosis and 25% of people with depression achieve full remission and response to pharmacological treatment, with the remainder achieving partial response or response without remission (S. Brown et al., 2010; Health (UK), 2014; Kern et al., 2020; Owen et al., 2016). Biologically-driven illness models, able to relate to those at highest risk of poor outcome and chronicity may allow new and targeted treatments to be delivered early (Linden, 2012). However, recognizing patients on a path to chronic disability, at an early stage, is still difficult in both psychosis and depression (Dinga et al., 2018; Klosterkötter et al., 2011). Previous transdiagnostic research has stressed the need for the use of machine learning (Fusar-Poli, Stringer, et al., 2019) and has identified specific patterns of

neurocircuit disruption across major psychiatric disorders in emotional reactivity and regulation (McTeague et al., 2020). Reininghaus and colleagues, building on previous calls for a dimensional approach to psychosis (Maj, 2016), have shown the use of multidimensional item response modelling to predict psychosis biotypes transcending traditional diagnostic boundaries; with suggestion of an underlying transdiagnostic dimension across psychotic diagnoses (Quattrone et al., 2019; Reininghaus et al., 2019; Tamminga et al., 2013). Recent semi-supervised machine learning studies using neuroanatomical data have identified the presence of an impaired neuroanatomical cluster which is characterized by overall poorer outcomes and functioning in schizophrenia (Chand et al., 2020) and in youth with internalizing symptoms (Kaczurkin et al., 2020). However, there has not yet been a transdiagnostic investigation of neuroanatomy specifically in depression and psychosis.

Herein, we aimed to identify replicable neuroanatomical clusters across patients with recent onset depression (ROD) and recent onset psychosis (ROP). We hypothesized that neuroanatomically derived clusters would be transdiagnostic, and related to distinct phenotypes drawn from symptom, neurocognitive, and inflammatory data across both disorders. We further aimed to explore the predictive validity of neuroanatomically identified clusters and externally validated our neuroanatomically based clusters in chronic depression and chronic schizophrenia, in an accelerated longitudinal design. We also developed supervised machine learning models to predict symptom remission in ROP and ROD and our neuroanatomically based transdiagnostic clusters. We hypothesised that models developed in neuroanatomically based transdiagnostic clusters will show greater predictive accuracy compared to those in traditional diagnostic groups.

Methods

Study design

This study utilizes data from the PRONIA study, an EU-FP7 funded seven-centre study as well as three external validation datasets. Details of the PRONIA study sites, recruitment protocol and quality control procedures can be found in the supplementary methods (1.1, 1.2, 1.3, tables S1, S2, S3) and a prior publication (Koutsouleris et al., 2018a). Data used in this analysis included structural MRI, demographic, clinical, neurocognitive and blood-based biomarker measures. See supplement for full details.

Inclusion and Exclusion Criteria

In brief, ROP participants had to meet the following criteria: 1) DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders, Text Revision) (American Psychiatric Association, 1994) affective or non-affective psychotic episode (lifetime), 2) criteria for DSM-IV-TR affective or nonaffective psychotic episode fulfilled within past 3 months and 3) onset of psychosis within past 24 months. ROD patients had to meet the following criteria: 1) DSM-IV-TR major depressive episode (lifetime), 2) major depressive disorder criteria fulfilled within past three months and 3) duration of first depressive episode no longer than 24 months. General inclusion criteria can be found in the supplement (1.5).

MRI imaging data acquisition, quality control, and preprocessing

Participants underwent a multi-modal MRI protocol. A minimal harmonization protocol, which the MR sequences across the different scanners had to comply with as well as the imaging preprocessing is described in the supplementary methods (1.3 and 1.4).

MRI harmonization and data acquisition

When setting up the PRONIA study, we decided to generate a MRI database that would represent the MR scanner sequence heterogeneity encountered in clinical real-world. The aim of this strategy was to strengthen the generalizability and clinical applicability of the predictive models developed by our machine learning analyses. Thus, we agreed on a minimal harmonization protocol that required the PRONIA sites to only (1) acquire isotropic or nearly isotropic voxel sizes of preferably 1 mm resolution, (2) set the Field Of View (FOV) parameters accordingly to guarantee the full 3D coverage of the brain including all parts of the cerebellum, and (3) define the relaxation time (TR) and echo time (TE) as well as other imaging parameters in a way that would maximize the contrast between cortical ribbon and the white matter and enhance the signal-to-noise ratio in the images. At every site all the images were visually inspected, automatically defaced and anonymized using an in-house Freesurfer-based script before the data was centralized.

MRI processing pipeline

The manual of the CAT12 toolbox (<http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>) details the processing steps applied to the structural images. These steps consist of:

- (1) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering(Manjón et al., 2008).
- (2) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local variations of intensity distributions as slowly varying spatial functions and thus achieves a homogeneous segmentation across cortical and subcortical structures(Rajapakse et al., 1997).

- (3) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior information of adjacent voxels into the segmentation estimation generated by AMAP(Rajapakse et al., 1997).
- (4) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM) inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP segmentation.
- (5) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the AMAP-generated tissue segments.
- (6) A high-dimensional DARTEL registration of the image to a MNI-template generated from the MRI data of 555 healthy controls in the IXI database (<http://www.braindevelopment.org>).
- (7) The GM maps were then multiplied with the Jacobian determinants that were obtained during registration in order to produce GM volume maps.
- (8) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

Semi-Supervised Machine Learning Analysis:

Heterogeneity through Discriminant Analysis (HYDRA) (Varol et al., 2017) is a semi-supervised machine learning clustering algorithm able to dissect disease heterogeneity by portioning patients based on patterns or transformations between the sub-populations (i.e., clusters) from the patient group and the reference group (i.e., healthy controls) through the use of a convex polytope formed by combination of multiple linear max-margin classifiers (i.e., support vector machines) and is able to regress out nuisance covariates, such as age and gender. We used the python version of HYDRA (<https://github.com/anbai106/pyHYDRA>)

(Varol et al., 2017) to simultaneously classify patients (ROP+ROD) from HC, and partition patients into clusters based on disease-related heterogeneity using structural MRI.

ComBat Harmonization

To mitigate site effects, prior to applying HYDRA, the R version of the ComBat harmonization technique was employed (<https://github.com/Jfortin1/ComBatHarmonization>). ComBat utilizes an empirical Bayesian framework that removes variance which is attributed to scanner differences while retaining disease effects. To further ensure that disease variance would be retained distinct from scanner variance, ComBat was trained on HC and then derived estimates were applied to the patients.

Model Training

We used whole volume (GMV and cerebrospinal fluid) brain measures derived from 280 regions of the neuromorphometrics atlas parcellation (CAT12) (four regions excluded due to zero variance) from 577 participants with ROP and ROD and HC (discovery sample of the PRONIA study). ROP patients and ROD patients were grouped together into one patient group. HYDRA was trained using a repeated hold-out cross-validation strategy (i.e., 1000 repetitions with 80% of the data for training in each repetition). Age, sex, and Total Intracranial Volume (TIV) were controlled as covariates. HYDRA was ran for 2 to 8 clustering solutions, and Adjusted Rand Index (ARI) was used to measure cluster stability. The most stable cluster solution was selected for further analysis. The statistical significance of clusters was assessed in three ways including testing our clustering solution against a gaussian distribution which assumes a dimensional severity explanation of our data. Details can be found in the supplement (1.11).

Phenotype Characterization

Identified clusters were compared to each other and to HC in terms of neurocognitive performance, blood-based biomarker (IL1ra, S100B, IL6, TNF α , CRP, TGF β , and BDNF) (see supplement 1.6) and symptom differences (PANSS, BDI, SANS) with univariate statistics corrected for multiple comparisons using false-discovery rate (FDR).

Neuroanatomical differences were examined using voxel-based morphometry (two sample t-test, SPM12), to identify the brain regions that the neuroanatomically derived clusters differed on. See supplementary material (section 1.14) for further granular investigation of clinical and inflammatory marker differences between clusters.

Independent and External Validation

To examine the generalizability of neuroanatomically based clusters we developed a SVM model, using the 280 features that our HYDRA model was trained on (Chand et al., 2020), to classify patients from the discovery sample into the identified clusters. This SVM was applied to the PRONIA-independent replication sample of ROP and ROD patients (N=404), collected at a different timescale from the discovery sample (May 2016 to February 2019). ComBat was trained on the replication HC and applied to the replication transdiagnostic patient group to mitigate site effects in the replication dataset. The SVM validation model that was trained on the discovery data was then applied to the replication data.

We externally validated the neuroanatomically based PRONIA clusters using the developed SVM model, in three MRI datasets of patients with chronic schizophrenia (Centre for Biomedical Research Excellence (COBRE) and Mind Clinical Imaging Consortium (MCIC) and chronic depression (Munich (MUC)) in an accelerated longitudinal design framework (see supplementary methods 1.9 and 1.10).

Predictive Utility

We trained SVM models using symptom and blood-based biomarker data to predict symptom recovery (as defined by a Global Assessment of Functioning-Symptom (GAF-S) score of ≥ 61) (*Recovery from Schizophrenia*, 2007) at 9 months. To assess the predictive utility within the neuroanatomically based clusters and within ROP and ROD groups we trained 4 different SVM models (one for each different diagnosis ROP/ROD/Cluster 1/Cluster 2) and compared their predictive accuracy in terms of area under the ROC curve, balanced accuracy, sensitivity and specificity. Details can be found in the supplement (1.8). A detailed figure of the analysis pipeline can be seen in figure 1.

Results:

Demographic Information

One hundred fifty-five participants with ROP, 147 patients with ROD, and 275 HC from the discovery sample were included in the HYDRA semi-supervised machine learning analysis. The mean age of the ROP group was 25.3 [SD 5.5], the mean age of the ROD group was 25.9 [SD 6.2]), and the mean age of the HC was 25.5 [SD 6.4]. The ROP group consisted of 96 male and 59 female patients, the ROD group had 66 male and 81 female patients, and the HC group had 107 male and 168 female participants. A summary of sociodemographic and clinical information is provided in table 1. Sociodemographic and clinical information for the PRONIA replication and external validation samples (COBRE, MCIC, and MUC) is provided in the supplement (1.9).

HYDRA Semi-Supervised Machine Learning Analysis

The optimal clustering solution was two transdiagnostic clusters (Cluster 1, n=153, 67 ROP, 86 ROD and Cluster 2, n=149, 88 ROP, 61 ROD, ARI: .618). Patients in cluster 1 had a mean

age of 26.2 [6.2] and the ones in Cluster 2 had a mean age of 24.9 [5.4]. There were 78 male and 75 female patients in cluster 1 and 84 male and 65 female patients in cluster 2. The two clusters did not differ in terms of age ($p=.071$), sex distribution ($p=.358$), total intracranial volume ($p=.144$), or medication exposure and differed in terms of original diagnosis distribution ($p=.008$). A sociodemographic and clinical description of the two clusters can be found in table 1.

Cluster Statistical Significance

The clusters were statistically significant 1) in terms of whether they would be different than if there was no disease related variability present ($p=.010$), 2) in terms of whether the disease structures were different ($p<.001$), and 3) in terms of whether the data could be better explained by a single Gaussian distribution ($p<.001$) suggesting that our data could not be explained in terms of a single Gaussian (continuous) distribution assuming a dimensional severity model. Details of the statistical significance tests can be found in the supplement (1.11).

Clinical Characteristics Associated with Neuroanatomically based clusters

Cluster 2 revealed a more severe symptom presentation compared to cluster 1 with significantly higher scores in the positive ($t(287)=-2.8$, $p=.020$), negative ($t(287)=-2.2$, $p=.040$), and general ($t(287)=-2.7$, $p=.010$) PANSS domains. Patients in cluster 2 had higher negative symptoms in SANS symptoms of affective flattening ($t(284)=-2.7$, $p=.010$), alogia ($t(282)=-3.0$, $p=.020$), and attention deficit ($t(255)=-2.2$, $p=.040$). Patients in Cluster 2 also showed worse functioning (Global Functioning-Role) ($t(291)=-2.3$, $p=.030$). There were no statistically significant differences between the two clusters in terms of neurocognition or blood-based biomarker data in univariate analysis. All p values have been fdr corrected. See supplement tables S5, S6, and S7. In supplementary multivariate SVM analysis our

neuroanatomically based clusters were separable using cognitive data (BAC: 56.6%, sensitivity: 57.5%, specificity: 55.7%, AUC: 0.58, $p=0.01$). Patients in cluster 2 mainly exhibited worse cognitive performance in a visual recognition and recall task (Rey–Osterrieth complex figure) and patients in cluster 1 mainly performed worse in verbal memory tasks (Rey Auditory Verbal Learning Test) (See supplementary figures S6, S7, and S8). The two clusters were also separable (BAC: 58.7%, sensitivity: 54.9%, specificity: 62.4%, AUC: 0.59, $p=0.01$) in blood-based biomarkers, with patients in cluster 2 having elevated levels of CRP and TNF α (See supplementary figures S9, S10, and S11).

VBM analysis of neuroanatomically based clusters

We conducted a VBM analysis for the purpose of demonstrating the brain regions that the two clusters differed in. Here, Cluster 2 exhibited widespread GMV loss compared to Cluster 1 and also compared to HC in areas including the Superior Temporal Gyrus, the Cingulate Gyrus, and the Thalamus among others. Cluster 1 revealed increased GMV compared to HC in cerebellar areas. These results can be seen in figure 2 and in the supplement (tables S7 and S8 and figure S2).

Independent and External Validation

In independent validation the two-cluster model showed generalisability in the PRONIA replication sample with patients classified into the two clusters in the replication sample showing similar clinical and neuroanatomical patterns to the ones from the discovery sample (supplement section 1.18). When externally applied to the MCIC and COBRE (chronic schizophrenia) and MUC (chronic depression), patients from datasets with a higher mean of age and/or longer duration of illness were, more often placed in Cluster 2 as indicated by negative decision scores. The effects of duration of illness and age were statistically significant, $F(2,278) = 27.88$, $p < .001$. Post hoc analyses using the Tukey HSD post hoc

criterion for significance indicated that the mean decision score was significantly lower in the MUC group compared to the MCIC ($p < .001$). Mean decision score differences between the MCIC and COBRE ($p = .078$) showed a trend towards statistical significance. The results can be seen in table 2.

Prognostic Validation

Within the neuroanatomically based clusters, stacking a blood-based biomarker (IL1ra, CRP, TNF α , BDNF, and TGF β) SVM model to a symptom data (baseline PANSS, BDI, and GAF-S individual item scores) SVM model (i.e., a combined model) increased accuracy for predicting symptomatic recovery at 9 months (GAF-S) with BAC of 71.2% for cluster 1 and 57.0% for cluster 2. This outperformed a similar stacked blood-based biomarker and symptom data SVM model predicting GAF-S in ROP and ROD groups (table 3). A Kruskal-Wallis H test showed that there is a statistically significant difference between the outer cross-validation folds (CV2) BAC of the different models $H(3) = 22.9$, $p < 0.001$. Post-hoc Mann-Whitney U test results can be found in the supplement (1.13).

Discussion

In this study, we identified two transdiagnostic clusters across psychosis and depression, using semi-supervised machine learning and neuroanatomical data in a large sample of recent onset depression and psychosis patients. Both clusters contained similar numbers of patients with depression and psychosis, however they were clinically distinct, with one cluster being characterized by more general and negative symptom loading and functional impairment, widespread GMV loss, (hereafter called the “impaired” cluster) and one cluster characterized by fewer symptoms, less GMV loss, and less functional impairment but more ‘core’ depressive symptomatology (hereafter called the “preserved” cluster). The neuroanatomically based clusters were generalizable to a replication sample and further externally validated in

three datasets of patients with chronic illness. Patients with chronic illness, with a higher duration of illness and mean age, were more likely to be classified into the impaired cluster. We were further able to demonstrate that SVM learning models using clinical and blood-based biomarker data to predict symptom remission at 9 months showed a higher accuracy in the neuroanatomically derived clusters compared to traditional diagnostic categories.

The precise aetiology of mental illnesses including psychosis and depression remains elusive despite decades of research, with a stagnation in advance of new pharmacological and psychotherapeutic treatments (Brand et al., 2015; Dean & Keshavan, 2017; Radua et al., 2018). Our results suggest that current diagnostic categories, particularly in early stages of illness, may mask transdiagnostic phenotypes which include an identifiable group with greater impairment and poorer chance of remission across disorders. In our impaired cluster, patients had reduced GMV in areas that have been identified as central to the disease processes of both schizophrenia and depression, such as the superior temporal gyrus, the anterior cingulate, the insula, and the thalamus (Gray et al., 2020; Kempton et al., 2011; Schmaal et al., 2017; van Erp et al., 2018). In our analysis, a significant number of patients with depression, who may be perceived as having a less severe illness and better prognostic outlook than patients with psychosis, were ascribed to the impaired phenotype, suggesting that they are on a path towards poor outcome. Conversely, a significant number of patients with psychosis were not assigned to the impaired group, and therefore potentially have an identifiable early signature of good prognosis, which was further indicated by the fact that predicting 9-month symptomatic outcomes in that group was more accurate than traditional diagnostic groupings.

Categorical diagnoses have survived because some individuals (specifically those with chronic established illness) do indeed fit within these nosological entities and more valid solutions remain elusive to date (Potuzak et al., 2012). However, within the scope of affective

and non-affective major psychiatric diseases, the Kraepelinian dichotomy of dementia praecox and manic-depressive psychosis has long been challenged. Studies have shown that our understanding of the clinical and neurobiological distinction between disorders may be particularly challenging during early phases of illness (Birchwood et al., 2005; Craddock & Owen, 2010; Lalouis et al., 2021; Upthegrove et al., 2017). The concept of affective disorders as a differential diagnosis for psychosis, particularly in the early years of illness is waning, with recent research suggesting a central and causal role for depression in the pathogenesis of psychosis and mutual biological underpinnings. This further challenges the distinction between affective and non-affective pathways to psychosis (Craddock & Owen, 2010; Herniman et al., 2021; Lalouis et al., 2021; Upthegrove et al., n.d.). Fischer and Carpenter (Fischer & Carpenter, 2009) suggest that reducing heterogeneity in syndromes is essential to decisively address the Kraepelinian dichotomy. Despite the fact that dementia praecox does not directly map to non-affective psychosis, the Verrücktheit (chronic non-affective psychoses) made distinct in Kraepelin's first Edition (1883) led to the (mis)understanding that schizophrenia was non-affective (Kendler, 2020). The impaired cluster which contains both patients with schizophrenia and depression has more cognitive symptoms and a brain signature that is identified in our chronic replication sample. Deficit schizophrenia is a concept introduced over 30 years ago to reduce clinical heterogeneity and suggests the existence of a homogeneous schizophrenia subtype with persistent trait negative symptoms (Carpenter et al., 1988). The impaired cluster we identified could be characterized as a transdiagnostic deficit cluster across depression and psychosis due to its higher load of negative symptoms, a previously proposed marker of the deficit syndrome across diagnoses (Peralta & Cuesta, 2003). Furthermore, our findings of greater GMV reduction in the impaired cluster corroborate previous research which identified temporal GMV reduction as a marker of very poor outcome (Mitelman & Buchsbaum, 2007). Our neuroanatomically

derived clusters contained both patients with depression and psychosis in recent onset, replicated in our independent PRONIA sample. This suggests lack of diagnostic hierarchy across depression and psychosis, and that some syndromes may hold equal weight in association with poor outcome regardless of relationship to diagnosis. These results add to the challenge of the separation between affective and non-affective psychoses with affective and psychotic diagnostic groups featuring in both clusters; corroborating previous studies which found that high affective symptom scores were equally common in patients with affective and non-affective psychosis and question the clinical validity of such a distinction (van Os et al., 2000).

Our results support the common biological susceptibility model of psychiatric disorders and suggest that the biological underpinnings of disease course, at least in depression and psychosis, may be related to transdiagnostic mechanisms, which are potentially hidden by current nosological systems. A similar transdiagnostic model has previously been reported in genomic research, which has shown a certain degree of overlap in the biological susceptibility to mental illness across mood and psychotic disorders; evidence of a transdiagnostic biological cause of major psychiatric disorders is evident with the identification of genetic variants that confer a transdiagnostic risk for bipolar, major depressive disorder, and schizophrenia related to the Major Histocompatibility Complex featuring in both schizophrenia and depression genome wide association studies (Li et al., 2019; Mokhtari & Lachman, 2016). Our finding that elevated pro-inflammatory cytokines add to predictive accuracy of poor outcome in an impaired phenotype suggest that this genomic immune influence may be ongoing in those on a path to poor outcomes.

Schizophrenia GMV deficits in the hippocampus, temporal gyrus, and cerebellum are associated with genetic factors such as SATB2, GABBR2, and CACNA1C (Luo et al., 2019).

A common genetic basis between risk for altered brain structure and neuro-psychiatric

disorders has been conferred by findings of risk variant enrichment associations with brain structural phenotypes across diagnoses (Mufford et al., 2017). Our results suggest a transdiagnostic cluster of GMV impairment suggestive of common biological underpinnings for poor outcome across depression and psychosis with potentially more valid structures than traditional diagnostic categories for use in predicting symptomatic remission.

Heterogeneity and co-morbidity may be especially pronounced in the early stages of these disorders; this creates diagnostic uncertainty and difficulties in predicting disease and treatment course (Baca-Garcia et al., 2007; Keshavan & Brady, 2011; Koutsouleris et al., 2015; Pope et al., 2013; Salvatore et al., 2013). Our results suggest that a bottom-up approach based on neurobiological data may be more reliable in the elucidation of patients with potential for greater impairment and offer a potential future solution for the diagnostic challenges of mental illness. Our external validation findings show that the impaired cluster potentially identifies patients who are on a path to chronic illness from early stages of illness, given that the majority of patients in the external validation sample with chronic illness fell into the same cluster as our impaired group. This has potentially significant clinical implications in terms of personalised treatment and focused recovery interventions. The fact that patients from chronic samples with a higher mean age and illness duration were more likely to be assigned to the impaired cluster could be an indication that our neuroanatomically based clusters identify an accelerated transdiagnostic brain aging effect in recent onset samples, corroborating previous brain age studies (L. K. M. Han et al., 2020; Schnack et al., 2016).

Strengths and Limitations

The present analysis includes several strengths including a large dataset with rich clinical, neurocognitive, biomarker, and imaging data from both recent onset psychosis and

depression groups, independent and external validation, as well as significance testing of our clustering solutions (e.g. by testing whether the data could be better explained by a Gaussian distribution which assumes a dimensional severity explanation of the data). Furthermore, the technique we used for the identification of subgroups (HYDRA), offers a solution to issues that are usually associated with clustering based on unsupervised machine learning models which are built on biological data such as the detection of groups that may reflect underlying nuisance variance such as age, gender, body type, and common ancestry (genetics) (Dinga et al., 2019). Nevertheless, our results should be interpreted with caution as there are certain limitations. Due to the nature of our recent onset sample and using a healthy control sample as a reference group in the semi-supervised model, there is a risk that the differences between the groups are not as marked as would be seen in more chronic cases. We addressed that limitation by performing permutation tests to robustly assess the significance of the identified clusters. Furthermore, our models were developed in recent onset patients with a significantly lower mean age than that of our external validation samples. We addressed that limitation by following a robust pipeline that removed the age and site effects while retaining the disease variance in the data. Although we developed an accelerated longitudinal design with the use of recent onset and chronic samples and had a 9 month follow-up for prediction of symptom remission, definitive findings would need large longitudinal datasets with repeated measures, such as functional outcome, over many years. Finally, we only used neuroanatomical features to parse neurobiological variance among complex clinical presentations. Psychiatric illness is not a single variable problem and we have addressed that by examining whether the brain-based clustering solution is reflected in the phenotypic, cognitive, and inflammatory levels. Future studies should consider using multiple biological measures and larger population-level data to encompass the pleiomorphic nature of clinical entities such as depression and psychosis.

Conclusions

Using semi-supervised machine learning, we were able to identify two neuroanatomically based transdiagnostic clusters. One cluster was characterized by an impaired functional and neurocognitive profile and greater symptomatic loading and GMV loss while the other cluster was characterized by a more preserved neuroanatomical and reduced symptom signature. Our distinct impaired cluster included patients with depression and psychosis and may provide insight into transdiagnostic aetiopathogenetic pathways of chronicity and poor outcome. The identified clusters have been derived in recent onset samples using structural MRI and could eventually lead to the development of MRI-based prediction and decision-making tools. In external validation, older patients with longer duration of schizophrenia and depression were assigned in the impaired cluster suggesting a potential identifiable transdiagnostic signature of chronicity and path to poor outcome at the early disease stages. Using clinical and blood-based biomarker data, we were able to predict symptomatic and functional remission more accurately in the derived clusters compared to traditional diagnostic groups. Whilst such challenge to current diagnostic structures will need significant further replication and longer follow-up, identifying a transdiagnostic signature of poor prognosis has the potential to aid new and targeted treatment strategies across early stages of mental disorder.

Sample Sizes, No.	ROP Group	ROD Group	t/χ^2	P Value	Cluster 1 (Preserved)	Cluster 2 (Impaired)	t/χ^2	P Value	HC	HC vs Impaired Cluster		HC vs Preserved Cluster	
										t/χ^2	P Value	t/χ^2	P Value
Original Diagnostic Group (ROP/ROD)	155	147			153 (67/86) 43.2%/58.5%	149 (88/61) 56.8%/41.5%	$\chi^2 = 7.04$	.008	275				
Age, Mean (SD)	25.3 (5.5)	25.9 (6.2)	$t = -.879$	.380	26.2 (6.2)	24.9 (5.4)	$t = 1.81$	.071	25.5 (6.4)	$t = .887$	.375	$t = -1.035$	.301
Sex (Male/Female)	96/59	66/81	$\chi^2 = 8.8$	.003	78/75	84/65	$\chi^2 = .88$	.358	107/168	$\chi^2 = 11.9$	.001	$\chi^2 = 5.8$	.016
Total Intracranial Volume, Mean (SD)	1531.6 (141.9)	1500.6 (144.3)	$t = 1.87$	.061	1504.6 (144.0)	1528.7 (142.8)	$t = -1.46$	.144	1518.5 (140.8)	$t = -.708$	.481	$t = .975$	.330
Medication, Mean Cumulative Sum (SD)					5122.7 (16501.2)	11191.7 (52988.6)	$t = -1.24$	.214					
CPZE					390.5 (1780.0)	173.9 (551.4)	$t = -1.32$	.187					
OLAE					3095.7 (10409.5)	2504.3 (7975.8)	$t = .510$	.610					
SSRIE					282.8 (1031.5)	578.6 (3625.2)	$t = -.888$	.375					
BENZOE													
SCID Diagnosis, No. (%)													
Schizophrenia	63 (40.6)	0 (0)			22 (14.4)	41 (27.5)							
Schizophreniform Disorder	12 (7.7)	0 (0)			3 (2.0)	9 (6.0)							
Schizoaffective Disorder	8 (5.2)	0 (0)			4 (2.6)	4 (2.7)							
Delusional Disorder	8 (5.2)	0 (0)			7 (4.6)	1 (0.7)							
Psychotic Disorder NOS	22 (14.2)	0 (0)			11 (7.2)	11 (7.4)							
Major Depressive Disorder	13 (8.4)	140 (95.2)			88 (57.5)	65 (43.6)							
Bipolar Disorder I	9 (5.8)	0 (0)			4 (2.6)	5 (3.4)							
Other	20 (12.9)	7 (4.8)			14 (9.1)	13 (8.7)							
PANSS Positive Mean (SD)	17.5 (6.3)	7.6 (1.2)	$t = 18.25$	<.001	11.5 (5.8)	13.1 (7.4)	$t = -2.83$	.02					
PANSS Negative Mean (SD)	16.4 (7.9)	12.2 (4.7)	$t = 5.43$	<.001	13.5 (6.3)	15.2 (7.2)	$t = -2.21$	.04					
PANSS General Mean (SD)	35.7 (11.6)	27.1 (6.5)	$t = 7.99$	<.001	29.8 (8.2)	33.0 (11.4)	$t = -2.71$	.01					

Table 1. Sample Sociodemographics. Sample Sizes, Participants per Study Site, Age, Sex, Total Intracranial Volume, Medication. (Abbreviations: ROP=Recent Onset Psychosis, ROD=Recent Onset Depression, HC=Healthy Controls, SD=Standard Deviation, CPZE=Chlorpromazine Equivalent, OLAE=Olanzapine Equivalent, SSRIE=Selective Serotonin Reuptake Inhibitor Equivalent, BENZOE=Benzodiazepine Equivalent, SCID=Structured Clinical Interview for DSM Disorders, NOS=Not Otherwise Specified, PANSS=Positive and Negative Symptom Scale)

Diagnosis	COBRE		MCIC		MUC	
	Schizophrenia		Schizophrenia		Depression	
Sample Size	71		107		103	
Age M (SD)	38.1 (13.9)		34.5 (11.1)		42.1 (11.9)	
Duration of Illness M (SD)	16.8 (12.9)		10.9 (10.9)		5.8 (7.7)	
Mean Decision Score	-.04 (.63)		.15 (.71)		-.47 (.48)	

Table 2. External validation results. Decisions scores reflect mean distance of patients from the hyperplane separating the two clusters. Positive decision scores indicate assignment to cluster 1 (preserved cluster) and negative decision scores indicate assignment to cluster 2 (impaired cluster). F (2,278) = 27.88, p<.001

	True Positive, No.	True Negative, No.	False Positive, No.	False Negative, No.	Correct Unremitted, %	Correct Classification Rate	Correct Remitted, %	Balanced Accuracy, %	Positive Predictive Value, %	Negative Predictive Value, %	Model P Value
Stacked ROP 9-month Model	20	33	19	29	40.8	63.5	52.1	51.3	53.2	0.56	0.38
Stacked ROD 9-month Model	53	11	13	26	67.1	45.8	56.5	80.3	29.7	0.54	0.17
Stacked Preserved Cluster 9-month Model	19	54	11	13	59.4	83.1	71.2	63.3	80.6	0.72	0.07
Stacked Impaired 9-month Model	35	25	16	31	53.0	61.0	57.0	68.6	44.6	0.58	0.18

Table 3. SVM models predicting 9-month GAF-S remission. H (3)=22.9, p<0.001.

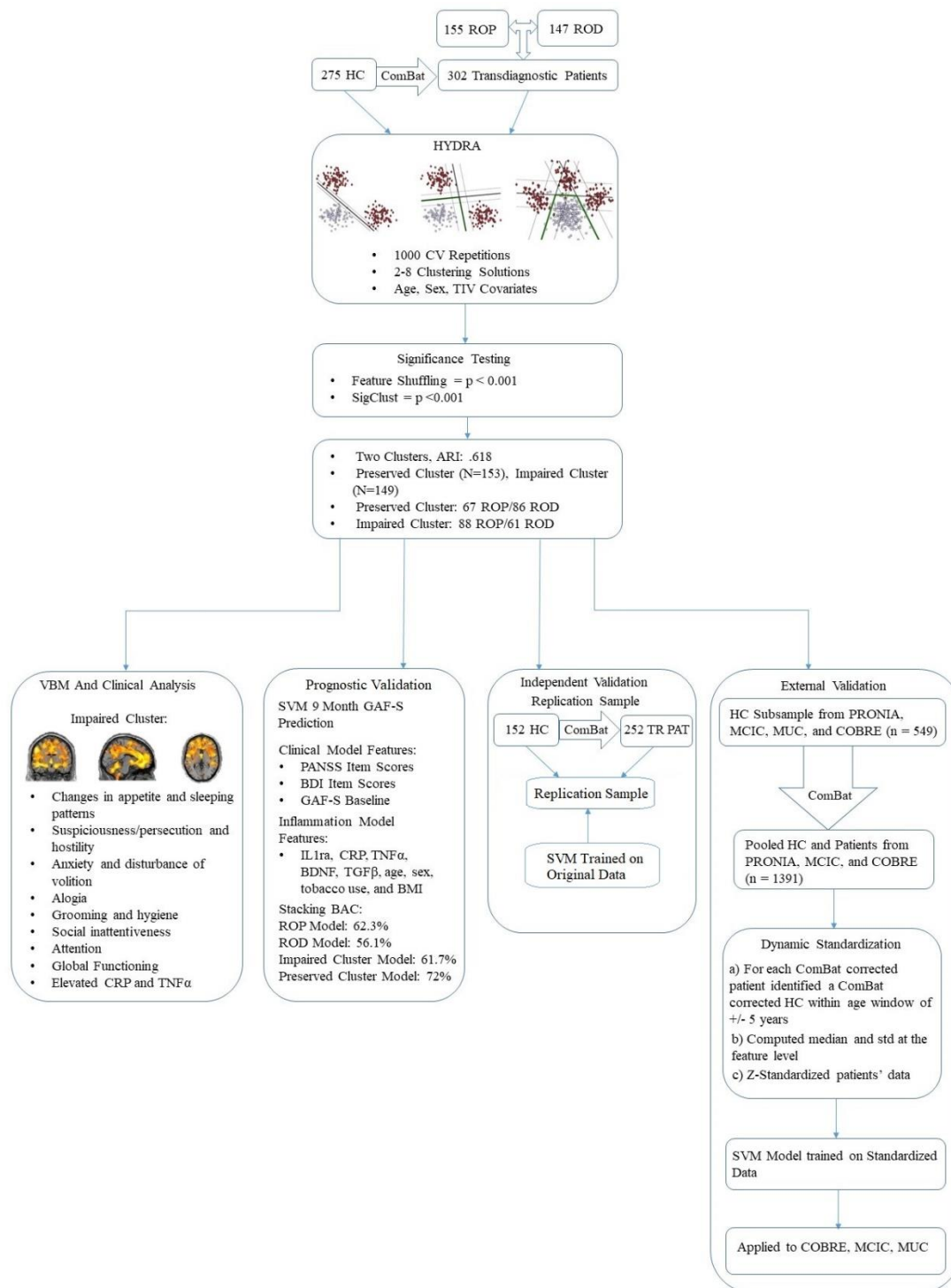


Figure 1: Analysis Pipeline Overview. This figure provides an overview of the analysis pipeline undertaken in this study. ROP and ROD patients were combined into one transdiagnostic group and ComBat was trained on HC and applied to the patients in order to remove site related variance from the data. The HC and the patient data were then entered

into the HYDRA algorithm with age, sex, and TIV, added as covariates. HYDRA was trained using a repeated hold-out cross-validation strategy (i.e., 1000 repetitions with 80% of the data for training in each repetition). The clusters were validated in the PRONIA replication sample and the three external datasets. Identified clusters were assessed for statistical significance and were then analyzed for clinical and VBM differences. Furthermore, the predictive utility of the clusters was assessed.

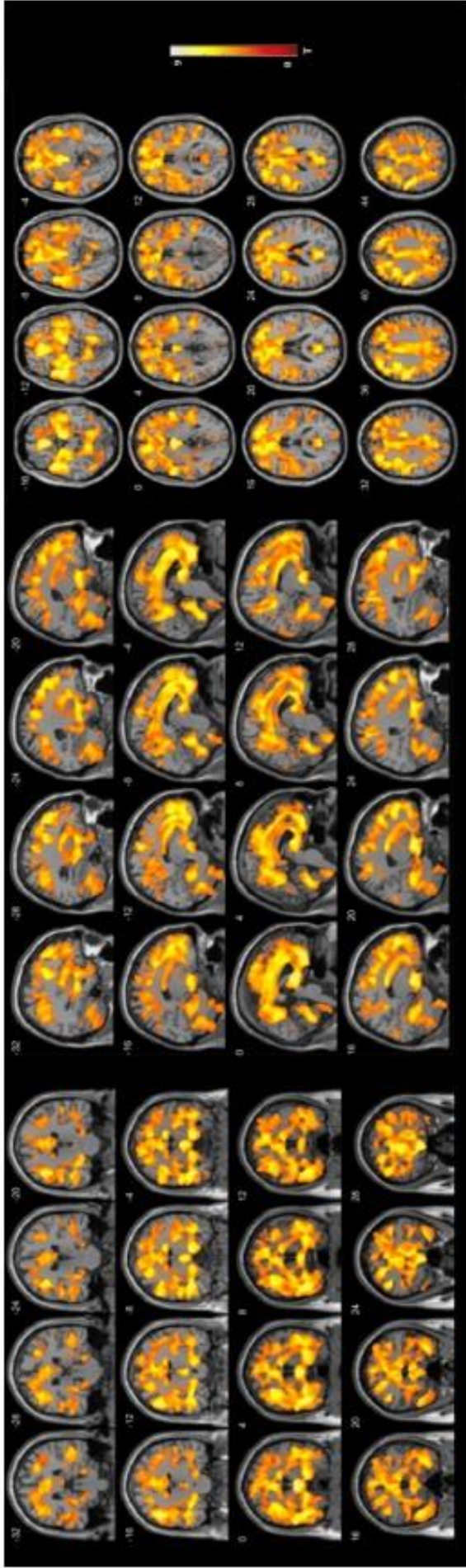


Figure 2: Impaired Cluster (Cluster 2) GMV Reductions Compared to the Preserved Cluster

(Cluster 1). GMV reductions are observed in the Middle Frontal Gyrus, Superior Frontal Gyrus, Superior Temporal Gyrus, Medial Frontal Gyrus, Cingulate Gyrus, Right Cerebellum, Left Cerebellum, Precuneus, Precentral Gyrus, Inferior Frontal Gyrus, Anterior Cingulate,

Chapter 6

Inflammatory Subgroups of Schizophrenia and their Association with Brain Structure: a Semi-Supervised Machine Learning Examination of Heterogeneity.

Paris Alexandros Lalousis, MSc^{1,2}; Lianne Schmaal, PhD^{3,4}; Stephen J. Wood, PhD^{1,3,4};
Renate L.E.P Reniers, PhD^{1,2,5}; Vanessa L. Cropley, PhD⁶; Andrew Watson, PhD⁷; Christos
Pantelis, MD, MRCPsych, FRANZCP^{6,8}; John Suckling, PhD⁹; Nicholas M. Barnes, PhD,
FBPhS¹⁰; Carmine Pariante, PhD¹¹; Peter B. Jones, MD⁹; Eileen Joyce, PhD, MRCP,
FRCPsych⁷; Thomas R.E. Barnes; DSc¹², Stephen M. Lawrie; MD, FRCPsych¹³, Nusrat
Husain, PhD¹⁴, Paola Dazzan, FRCPsych, PhD¹⁵; Bill Deakin, FRCPsych FMedSci¹⁶;
Cynthia Shannon Weickert, PhD^{17,18,19}; Rachel Upthegrove, FRCPsych, PhD^{1,2,20}

Author Affiliations:

1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United
Kingdom

3 -Orygen, Parkville, Australia

4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

5 -Institute of Clinical Sciences, University of Birmingham, United Kingdom

6 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia

7 -The Department of Clinical and Motor Neuroscience, UCL Queen Square Institute of
Neurology, London, UK

8 -NorthWestern Mental Health, Western Hospital Sunshine, St. Albans, Victoria, Australia

9 - Brain Mapping Unit, Department of Psychiatry, Herchel Smith Building for Brain and
Mind Sciences, University of Cambridge, UK; Cambridgeshire & Peterborough NHS
Foundation Trust, Cambridge, UK.

10 -Institute for Clinical Sciences, University of Birmingham, Birmingham, United Kingdom

11 - Stress, Psychiatry and Immunology Lab & Perinatal Psychiatry, The Maurice Wohl

Clinical Neuroscience Institute, King's College London, London, UK

12- Division of Psychiatry, Imperial College London, London UK

13- Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh,
Edinburgh, UK

14 - Division of Psychology and Mental Health, University of Manchester & Mersey Care
NHS Foundation Trust

15 -Department of Psychological Medicine, Institute of Psychiatry, Psychology and
Neuroscience, King's College London, London, UK

16 -Division of Neuroscience and Experimental Psychology, University of Manchester,
Manchester, United Kingdom

17 -Department of Neuroscience and Physiology, Upstate Medical University, Syracuse, NY,
USA

18 -Schizophrenia Research Laboratory, Neuroscience Research Australia, Sydney, New
South Wales, Australia

19 -School of Psychiatry, University of New South Wales, Sydney, New South Wales,
Australia

20 -Birmingham Early Interventions Service, Birmingham Women's and Children's NHS
Foundation Trust

Corresponding Author: Paris Alexandros Lalousis, BSc, MSc, School of Psychology,
University of Birmingham, 52 Pritchatts Road, Birmingham, B15 2SA, United Kingdom

()

Research in context

Evidence before this study There have been inconsistent findings in the identification of inflammatory subgroups of schizophrenia, and this may be the result of using traditional statistics or clustering methods which are vulnerable to underlying variance of age, gender, body type etc and fail to capture the potential heterogeneity within inflammation in schizophrenia. We searched PubMed with the terms “cluster” “inflammation” “schizophrenia” and “heterogeneity” on 16/02/2022. We found four studies in English which examined subgroups of schizophrenia and markers of inflammation. One study used inflammatory markers to distinguish individuals with schizophrenia and bipolar disorder. Another study analysed induced pluripotent stem cells carrying 22q11.2 deletion and examined inflammation markers in post-mortem brains. One review discussed the need to reduce heterogeneity in pathophysiological mechanisms such as inflammation in order to construct successful aetiopathogenetic models. Finally, a randomised controlled trial identified a subtype of schizophrenia associated with a specific inflammatory signature that was associated with symptomatological severity and higher odds of non-remission.

Added value of this study To our knowledge this is the first study to apply semi-supervised machine learning clustering to inflammatory serum data in patients with schizophrenia, with a large sample and exploration in an external dataset. We demonstrated multiple inflammatory clusters with distinct neuroanatomical profiles: 1) Low inflammation, 2) elevated CRP, 3) elevated IL6/IL8, 4) elevated INF- γ and 5) elevated IL-10. The IL6/8 cluster had the most widespread grey matter volume loss and greatest cognitive deficit. The IFN- γ cluster was the most structurally and functionally preserved. The Low inflammation and CRP clusters were identified in our external, younger sample.

Implications of all the available evidence We confirm that elevated cytokines are found in large portion of people with schizophrenia and that multiple subsets of inflammatory groups within schizophrenia can be distinguished. There may be multiple potentially causal inflammatory pathways relevant to schizophrenia with an inflammatory cluster characterized by increased CRP levels more evident in early years of illness and a cluster characterized by increased IL6/IL8 levels found in chronic patients, which is linked to structural brain deficits. This offers potential stage specific targets for further experimental medicine study and novel treatments.

Abstract

Introduction: Immune system dysfunction is implicated in the aetiology of schizophrenia, and may be related to gray matter volume (GMV) reduction. However not all patients with schizophrenia show evidence of inflammation. Attempts to identify an immune active subgroup using conventional clustering methods in relatively small samples have produced inconsistent results. We aimed to use an advanced clustering technique to examine the heterogeneity of inflammatory groups within a large sample of patients with schizophrenia.

Method: HYDRA (Heterogeneity through Discriminant Analysis) was used to separate patients with chronic schizophrenia (n=467) from healthy controls (HCs) (n=600) and define disease-related subgroups based on eight serum cytokines and c-reactive protein (CRP), in the Australia Schizophrenia Research Bank (ASRB). We tested GMV alterations and neurocognitive deficits in identified clusters. The Benefits of Minocycline on Negative Symptoms of Schizophrenia (BeneMin) sample was used to test whether identified clusters could be detected in a younger sample.

Outcomes: An optimal clustering solution revealed five main groups (Low Inflammation, elevated CRP, elevated IL-6/IL-8, elevated IFN- γ , and elevated IL-10) with an adjusted Rand index of 0.573. When compared with the healthy controls, the IL-6/IL-8 cluster showed the most widespread, including anterior cingulate GMV reduction. The IFN- γ inflammation cluster showed the least GMV reduction and impairment of cognitive performance. The two largest clusters (CRP and Low inflammation) were also identified in the younger BeneMin sample.

Interpretation: We identified five clusters of schizophrenia with distinct immune signatures, associated with differing GMV and neurocognitive function. The identification of

inflammatory subgroups has the potential to inform the development of novel, targeted treatments in schizophrenia.

Funding: National Institute for Health Research Efficacy and Mechanism Evaluation (EME) programme, an MRC and NIHR partnership, MRC MR/S037675/1. The Australian Schizophrenia Research Bank (ASRB), was supported by the Australian National Health and Medical Research Council (NHMRC) Enabling Grant (ID 386500), the Pratt Foundation, Ramsay Health Care, and the Viertel Charitable Foundation. Chief Investigators for ASRB were Carr, V., Schall, U., Scott, R., Jablensky, A., Mowry, B., Michie, P., Catts, S., Henskens, F., Pantelis, C. CP was supported by a NHMRC Senior Principal Research Fellowship (1105825), and NHMRC L3 Investigator Grant (1196508). CSW was supported by NHMRC L3 Investigator Grant (ID: 2009237).

Introduction

Immune system dysfunction is implicated in the aetiology of schizophrenia, with circulating cytokines such as interleukin (IL)-6, IL-1 β , interferon gamma (IFN)- γ , tumor necrosis factor alpha (TNF)- α , IL-12 and transforming growth factor beta (TGF)- β detected at elevated levels before the onset of psychosis and before treatment initiation (Khandaker et al., 2015; B. J. Miller et al., 2011; Noto et al., 2015; Perkins et al., 2015; Upthegrove et al., 2014).

Additionally, longitudinal studies have reported an association between higher adolescent C-reactive protein (CRP) levels and later development of schizophrenia (Metcalf et al., 2017) with a causal role of immune dysfunction suggested by genetically determined levels of IL-6 associated with schizophrenia using mendelian randomisation (Perry et al., 2021) and copy number variants in C4 associated with increased risk of schizophrenia (Sekar et al., 2016). Genome wide association studies also show evidence of risk for schizophrenia related to the Major Histocompatibility Complex (MHC) (Li et al., 2019; Mokhtari & Lachman, 2016).

Peripheral inflammation, as measured by cytokines and acute phase proteins, has been associated with reduced gray matter volume (GMV) in the hippocampus, orbital frontal cortex, inferior frontal gyrus, middle frontal gyrus, and cingulate cortex in people with schizophrenia (Fillman et al., 2016; B. J. Miller et al., 2014; North et al., 2021; Wu et al., 2019). Cortical thickness may be differentially affected by immune changes, with total cortical volume negatively associated with elevated IFN- γ and IL-5 in patients with first episode of psychosis (FEP) (Laskaris et al., 2021). Furthermore, elevated CRP and pro-inflammatory cytokines have been associated with significant cognitive impairment in patients with schizophrenia (Bulzacka et al., 2016; Fillman et al., 2016; Johnsen et al., 2016; Misiak et al., 2018; North et al., 2021).

It is likely that increased inflammation may be found in only a subgroup of patients with schizophrenia. However, the current evidence is mixed; Pillinger et al. found no evidence of a bimodal variation in cytokines in first episode schizophrenia, proposing that alterations in the immune system may be a core feature across all psychoses (Pillinger et al., 2019). Fillman et al. demonstrated the existence of two (elevated/low pro-inflammatory cytokine transcript) inflammatory schizophrenia subgroups with the elevated cytokine group having worse verbal fluency and reduced GMV in Broca's area but no differences in clinical symptoms (Fillman et al., 2016). Boerrigter et al. identified three (very elevated/elevated/low pro-inflammatory cytokine) inflammatory subgroups in people with chronic schizophrenia (Boerrigter et al., 2017) and Tamouza et al. reported two cytokine subgroups; (elevated/low pro-inflammatory) in patients with schizophrenia (Tamouza et al., 2021).

The inconsistency of findings may result in part from variation in patient samples, or with multiple inflammatory mechanisms operating at different stages of illness, which are not currently revealed by traditional analytical approaches and/or small sample sizes that result in reporting broad quantitative differences (e.g. low vs high). The use of advanced statistical methods including clustering, in larger samples is needed to identify subgroups characterised by qualitatively distinct inflammatory profiles.

Most clustering studies in this field to date have employed two-step hierarchical clustering with relatively small samples and few have performed external validation (Boerrigter et al., 2017; Fillman et al., 2016; Tamouza et al., 2021). Advanced clustering analysis methods have the potential to uncover inflammation subgroups in schizophrenia which may reflect differential pathological processes and aid the stratification of novel treatment targets (Varol et al., 2017). We used the novel semi-supervised machine learning tool, Heterogeneity through Discriminant Analysis (HYDRA) (Varol et al., 2017) with the aim of identifying subgroups of patients with schizophrenia characterised by distinct inflammatory profiles, and

investigating their neuroanatomical and neurocognitive correlates. In contrast to non-linear kernel classification methods which do not provide explicit information on disease subtypes HYDRA captures multiple dimension of heterogeneity by varying numbers of hyperplanes. Each hyperplane separates healthy controls from patients and then creates between-patient clusters, providing information about the distinct disease processes that drive the heterogeneity of the data.

We hypothesized that multiple inflammatory subgroups would be identified, and that those groups with an elevated inflammatory profile would have less GMV and poorer cognitive function. We aimed to extend our clustering solution in an external data set of participants at an earlier stage in schizophrenia, to reveal potential stage-specific insights.

Methods

Participants

This study used data from the Australia Schizophrenia Research Bank (ASRB) comprising 1067 participants: 467 patients with chronic schizophrenia and 600 healthy controls (HCs). The exclusion criteria were organic brain disorder, brain injury with greater than 24 hours post-traumatic amnesia, learning disability (IQ <70), movement disorders, current diagnosis of substance dependence, electroconvulsive therapy received in the last 6 months and, for controls, a personal or family history of psychosis or bipolar 1 disorder.

Blood-Based Biomarker Data

A 40 mL whole blood sample was collected drawn under non-fasting conditions at various times and collected into Serum Separator Tubes (SST, BD Biosciences). The tubes were inverted and left at room temperature for 30 minutes (to allow clotting) and centrifuged at 2000g for 5 minutes at 4°C. In the BeneMin dataset plasma inflammatory markers were aliquoted and frozen within 4 h. See previous publications(Loughland et al., 2010; North et

al., 2021) and supplementary methods for further details and BeneMin blood-based biomarker sampling and data analysis (1.1).

MRI imaging data acquisition, quality control, and preprocessing

Structural magnetic resonance imaging (sMRI) data were available for 296 participants: 195 patients and 101 HCs. sMRI brain scans were performed using a standard data acquisition protocol on Siemens Avanto 1.5 Tesla scanners located across Australia. See supplementary methods for full details of acquisition (1.2).

The T1-weighted (MPRAGE) images consisted of 176 1 mm thick sagittal slices per brain without gap; field of view = 250° - 250 mm; repetition time/echo time = 1980/4.3 ms; data matrix size = 256°-256; voxel dimensions = 1.0°-1.0°-1.0 mm. All scans underwent quality control and those with movement artefacts and/or brain lesions were excluded. The images were processed using the open-source CAT12 toolbox (version r1155; <http://dbm.neuro.uni.jena.de/cat12/>) (see supplementary methods 1.2). Whole-brain GMV measures were used.

The details of the MRI processing steps can be found in the CAT12 toolbox manual (<http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>) In short these steps consist of:

- (1) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering.
- (2) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local variations of intensity distributions as slowly varying spatial functions and thus achieves a homogeneous segmentation across cortical and subcortical structures.

- (3) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior information of adjacent voxels into the segmentation estimation generated by AMAP.
- (4) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM) inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP segmentation.
- (5) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the AMAP-generated tissue segments.
- (6) A high-dimensional DARTEL registration of the image to a MNI-template generated from the MRI data of 555 healthy controls in the IXI database (<http://www.braindevelopment.org>).
- (7) The GM maps were then multiplied with the Jacobian determinants that were obtained during registration in order to produce GM volume maps.
- (8) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

Neurocognition

Premorbid IQ was assessed using The Wechsler Test of Adult Reading (WTAR)(Kreutzer et al., 2011). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)(Randolph et al., 1998) and the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III) Letter Number Sequencing (LNS) subtest (age-adjusted scaled scores)(Kreutzer

et al., 2011) were used to examine a range of neurocognitive performance. Patients with missing data were excluded from statistical analyses.

External Dataset:

We used the baseline cytokine, clinical, cognitive and demographic data from 199 patients who had participated in the Benefit of Minocycline on Negative Symptoms of Psychosis: Extent and Mechanism (BeneMin) trial for exploration of our clusters in an external dataset. BeneMin is described in full in (Deakin et al., 2018) but in brief, the participants were within 5 years of onset of first episode of schizophrenia, schizophreniform, or schizoaffective psychosis. See supplementary methods for full details (1.1).

Semi-Supervised Machine Learning Analysis:

We used the python version of HYDRA (<https://github.com/anbai106/pyHYDRA>) (Varol et al., 2017) in the ASRB data to simultaneously classify patients from the HCs, and partition patients into clusters based on disease-related heterogeneity using CRP, IFN- γ , IL-10, IL-12, IL-1- β , IL-2, IL-6, IL-8, and TNF α . HYDRA (Varol et al., 2017) utilizes a convex polytope formed by combination of multiple linear max-margin classifiers; i.e. multiple support vector machines (SVM). It dissects disease heterogeneity by portioning patients into groups based on the patterns or transformations between sub-populations (i.e., clusters) in the patient group *as distinct from* the reference group (i.e., healthy controls). Furthermore, HYDRA models regress out entered covariates, such as age and sex.

Model Training

We used the nine aforementioned cytokines/acute phase response protein from 1067 participants (467 patients and 600 HCs) in the ASRB data set. HYDRA was trained using a repeated hold-out cross-validation strategy (i.e., 1000 repetitions with 80% of the data for

training in each repetition). Age, sex, site, and time the samples spent in the freezer were included as covariates. HYDRA was run requesting 2 to 8 clustering solutions, and the Adjusted Rand Index (ARI) was used to measure cluster stability.

The statistical significance of clusters was assessed in two ways. First, we assessed whether the obtained subgroups were different from those which would be found if disease-related variability was not present by comparing the observed ARI to a null distribution of ARI. We created a null distribution of ARI by randomly assigning HC samples (n=600) to an HC group and pseudo-patient group and performing a HYDRA analysis 100 times using the aforementioned repeated hold-out cross-validation strategy. We then compared the two distributions using a two-sample t-test. Secondly, we also assessed whether the data could be better explained by a single Gaussian distribution by employing SigClust (1000 simulations)(Huang et al., 2015) (<https://github.com/pkimes/sigclust2>). SigClust tests the null hypothesis that the data are from a single Gaussian distribution by using a test statistic called the cluster index (ClustI). The ClustI is defined as the sum of within-class sums of squares about the mean divided by the total sum of squares about the overall mean. The null distribution of the ClustI is approximated by simulating from a single Gaussian distribution estimated from the data.

Cluster Characterization

To aid the description of the clusters resulting from the HYDRA output we tested if inflammatory and neurocognitive profiles differed according to subgroup, and we interrogated confounds that were regressed out in the HYDRA model (age, sex, site, and time the samples spent in the freezer). Identified clusters were compared in terms of their cytokine levels, demographics, and neurocognitive performance using one-way ANOVA models corrected for multiple comparisons (Tukey's HSD).

Neuroanatomical differences were examined using voxel-based morphometry (VBM) (two sample t-tests of each cluster vs the HCs and a voxel-wise F test between identified clusters without the HCs, with age, sex, TIV, and site included as covariates, SPM12). All models were false discovery rate (FDR) corrected for multiple comparisons and a minimum cluster extent of $k=30$ was applied.

Cluster Solution Generalizability

In order to be able to test the generalisability of the cluster solution to the BeneMin dataset, first we developed a SVM model to predict subgroup membership in the ASRB dataset. The resulting SVM model was then applied to the BeneMin dataset. Ensemble-based probability (EBP), a measure of the probability an individual belongs to a cluster, was used to test the probability of a BeneMin participant falling into one of the ASRB identified clusters. Further details about the SVM parameters can be found in the supplement (1.4). Identified clusters were compared on demographic data, length of illness and neurocognitive performance using one-way ANOVA models.

Role of the funding source

The funders had no role in study design, data collection, data analysis / interpretation, or report writing. The corresponding and senior authors had full access to study data. The corresponding author had final responsibility for the decision to submit for publication.

Results

Demographic Information

The mean age of the ASRB HC group was 43.5 years [SD 13.1] and the mean age of the patient group was 39.5 years [SD 10.9]. The HC group consisted of 240 male and 360 female patients and the patient group had 303 male and 164 female participants. A summary of

sociodemographic and clinical information is provided in table 1. The mean age of the patients in the BeneMin external dataset was 25.6 years [SD 5.2]. One hundred and forty six patients were male and 53 were female.

HYDRA Semi-Supervised Machine Learning Analysis

The optimal clustering solution was eight clusters with an ARI 0.573. Due to very small number of patients in clusters 6-8, which would prohibit any meaningful conclusions, they were omitted from further analyses. Clusters 1-5 did not differ in terms of age ($p=.611$) or site distribution ($p=.358$), but did differ in terms of time the samples spent in the freezer (Cluster 1<Cluster 2; $p=.002$) and sex distribution, with more females in HCs ($p=.021$). Sociodemographic and full cytokine profiles of each clusters can be found in Table 1.

Cluster Statistical Significance

The clustering solution was significantly different from a clustering solution derived from HCs ($p<.001$) and the subgroup solution was more likely than an underlying Gaussian solution ($p<.001$). This means that the clusters identified do not exist in HCs and are specifically related to disease related inflammation and that the data distribution is not unimodal but rather that separable clusters exist within the dataset. See supplement (1.5).

Inflammatory Characteristics Analysis

For clusters 1-5 we assigned labels that depicted the most notable inflammatory changes in each group (see Table 1). Cluster 1 (Low Inflammation $n=142$) was characterised by statistically significant lower levels of cytokines and CRP (CRP, IFN- γ , IL-10, IL-12p70, IL-2, IL-6, and IL-8 respectively), compared with the HC group ($F(5, 1056) = 133.3, p<.001$; $F = 33.8, p<.001$; $F = 20.3, p<.001$; $F = 23.1, p<.001$; $F = 12.8, p<.001$; $F = 16.5, p<.001$; $F = 113.9, p<.001$) and the rest of the patient clusters ($F(4, 456) = 151.8, p<.001$; $F = 63.9,$

$p < .001$; $F = 28.7$, $p < .001$; $F = 27.5$, $p < .001$; $F = 22.3$, $p < .001$; $F = 22.3$, $p < .001$; $F = 193.1$, $p < .001$). Cluster 2 (CRP cluster; $n=121$) had significantly elevated levels of CRP compared with the rest of the clusters ($F(4, 456) = 151.8$, $p < .001$) and the HCs ($F(5, 1056) = 133.3$, $p < .001$). Cluster 3 (Classic Inflammation cluster; $n=82$) demonstrated elevated levels of IL-6 and IL-8 compared with the HCs ($F(5, 1056) = 16.5$, $p < .001$; $F = 113.9$, $p < .001$) and all the other patient clusters ($F(4, 456) = 22.3$, $p < .001$; $F = 193.1$, $p < .001$). Cluster 4 (IFN- γ cluster; $n=80$) exhibited significantly elevated IFN- γ levels compared with the HCs ($F(5, 1056) = 33.8$, $p < .001$ and all other patient clusters ($F(4, 456) = 63.9$, $p < .001$). Cluster 5 (Anti-Inflammatory cluster; $n=32$) exhibited elevated IL-10 levels compared with the HCs ($F(5, 1056) = 20.3$, $p < .001$ and all other patient clusters ($F(4, 456) = 28.7$, $p < .001$). See Table 1 and supplement 1.6 for post-hoc comparisons.

VBM analysis of neuroanatomically based clusters

Within the diagnosis of schizophrenia, Clusters 1-5 all had temporal and hippocampal GMV reduction compared to HCs. The Classic Inflammation cluster (Cluster 3) showed the greatest and most widespread GMV reduction, including the anterior cingulate. The IFN- γ cluster (Cluster 4) exhibited GMV increases in the precentral gyrus. See Figures 1-3, supplementary Tables 1-5. When all inflammatory clusters were compared with each other, decreases in the inferior temporal gyrus, dorsolateral superior frontal gyrus, and postcentral gyrus were observed in the CRP and Classic Inflammation clusters. See supplement 1.7 for further details.

Neurocognitive differences

Univariate ANOVAs for each neurocognitive domain (RBANS immediate memory/constructional/language/attention/delayed memory/total domains, WTAR, WASI matrix reasoning, WASI IQ equivalent, LNS respectively) showed a main effect of group in

all domains ($F(5,1046) = 100.2$; $F = 22.7$; $F = 56.7$; $F = 98.6$; $F = 81.3$; $F = 138.6$; $F = 22.0$ $F = 52.2$; $F = 50.1$; $F = 70.0$; $F = 54$, all p values $<.001$). Post-hoc analyses using the Tukey HSD post-hoc criterion ($\alpha=.05$) for significance showed that patients in all clusters had significantly lower scores in all domains compared with the HCs, apart from the premorbid functioning domain where differences between patients in the IFN- γ cluster and HCs showed a trend towards statistical significance ($p=0.076$).

Repeating the univariate ANOVAs without the HCs, in order to assess differences between schizophrenia clusters on neurocognitive performance, we observed a trend towards a significant main effect of group in the following domains: premorbid functioning ($F(4,446) = 2.12$, $p=.077$) and verbal memory (LNS) $F(4,447) = 2.16$, $p=.072$.

The Low and Anti-Inflammatory clusters both had low/normative cytokine and CRP levels and could be considered to both be part of the same low cytokine group, thus were combined to increase statistical power for the purposes of comparing the neurocognitive performance tests with the inflammatory clusters 2-4. Main effects of cluster were observed in the following neurocognitive domains: language processing (RBANS language), $F(3,447) = 3.92$, $p=.009$; attention (RBANS attention), $F(3,446) = 3.73$, $p=.011$; verbal working memory (LNS), $F(3,448) = 2.87$, $p=.036$, and premorbid functioning (WTAR), $F(3,446) = 2.76$, $p=.041$. Post-hoc comparisons are given in the supplement (1.7). Overall, the IFN- γ cluster showed the least neurocognitive dysfunction compared with the rest of the clusters.

External Dataset Cluster Exploration

Two main clusters were detected in the BeneMin dataset: Low inflammation cluster ($n=149$; (mean EBP: .95 [SD .00]) and the CRP cluster ($n=48$; mean EBP: .98, [SD .00]). The anti-inflammatory cluster ($n=2$; mean EBP: .97 [SD .02]) and the IFN- γ cluster ($n=1$; mean EBP: .99 [SD N/A]) were also identified but restricted to one or two patients. The classic

inflammation cluster was not detected, possibly due to lower IL-6 and IL-8 mean values in BeneMin. Patients in the CRP cluster exhibited significantly lower current IQ ($p = .028$) and premorbid IQ ($p = .005$), and worse performance on the Digit Symbol Substitution task ($p = .003$) compared with the Low Inflammation cluster. Further details can be found in the supplement (1.9).

Discussion

Using advanced semi-supervised machine-learning, we identified five distinct inflammatory subgroups of chronic schizophrenia. The clustering solution was statistically significant by two tests 1) the solution was unique to disease related variability 2) the data could not be better explained as having a continuous distribution. GMV loss in temporal and hippocampal areas was evident in all schizophrenia clusters when compared with the HCs, and the Classic Inflammation cluster, with elevations in both IL6 and IL8, showed the most widespread loss extending into the anterior cingulate. Greater GMV in the precentral gyrus was seen only in the IFN- γ cluster, the cluster which also had the lowest CRP and IL-6 values. The CRP and Low Inflammation clusters were detected in patients in BeneMin, suggesting that CRP could be a marker used to discriminate inflammatory subgroups at an earlier disease stage. This supports previous research suggesting that CRP is a potential biomarker of later disease onset(Perry et al., 2021).

Each identified cluster had an overlapping and distinct inflammatory *and* neuroanatomical profile. The CRP cluster showed increased CRP alone, and an associated pattern of reduced GMV in the inferior temporal gyrus, dorsolateral superior frontal gyrus, and postcentral gyrus. A potential mechanism for the effects of CRP on GMV could be that increased CRP levels (indicative of systemic low-grade inflammation, but not necessarily simultaneous with elevations in cytokines- as CRP is a more stable marker than cytokines)(Jacomb et al., 2018)

might result in an increase of blood–brain barrier permeability and subsequent pro-inflammatory cytokine entrance to the central nervous system. However, it seems unlikely that mild elevations in CRP would be a sufficient proximal cause of the GMV loss. More plausibly, raised CRP and reduced GMV could independently reflect an earlier developmental or ongoing neuroimmune abnormality such as maternal immune activation (MIA) or be a sign of more chronic inflammation. The relationship of higher CRP and altered neurodevelopment is supported by findings of a separate analysis of BeneMin participants which found that a subgroup of patients with low premorbid and low current IQ had higher levels of CRP and smaller MRI-derived total brain and intracranial volumes when compared to subgroups with a normal premorbid IQ(Watson et al., 2022). As these MRI findings are indicative of smaller premorbid brain size, this was suggestive of a neurodevelopmental subtype of schizophrenia additionally characterised by impaired premorbid cognitive function and low-grade inflammation(Watson et al., 2022). Nevertheless, research in rodents has shown that MIA does not always alter peripheral cytokines in male or female offspring(A. M. Brown et al., 2022).

Patients in the Classic Inflammation cluster had higher IL-6 and IL-8 levels compared with all other clusters and HCs. In the brain, IL-6 and IL-8 transcripts are extremely elevated in the frontal cortex in about 40% of those with schizophrenia(Fillman et al., 2013, 2016; Y. Zhang et al., 2016b). Peripheral IL-6 is the most consistently increased cytokine in all phases of schizophrenia(Fond et al., 2020) and the most consistently elevated cytokine found in the brain of people with schizophrenia with increased IL-6 found in the dorsolateral prefrontal, orbital frontal, midbrain and subependymal zone(Fillman et al., 2013, 2016; North et al., 2021; Purves-Tyson et al., 2020; Y. Zhang et al., 2016b). Patients in the Classic Inflammation cluster exhibited the most widespread GMV loss compared with HCs and other clusters, with specific reduction in anterior cingulate cortex. Anterior cingulate GMV may be an

endophenotypic biomarker resulting from reductions in neuronal, synaptic, and dendritic density (Fornito, Yucel, et al., 2009). Our finding of anterior cingulate GMV reduction in the Classic Inflammation cluster may suggest the potential for IL-6 function as a key mediator of GMV loss with relevance to schizophrenia. Recent evidence from Mendelian Randomisation also demonstrated potential causality of circulating levels of IL6/IL6R and GMV loss relevant to neuropsychiatric disorders including schizophrenia (J. Williams et al., 2022). Inhibition of IL-6 function in targeted groups of patients with schizophrenia holds significant promise not only for the alleviation of positive symptoms but also ameliorating cognitive impairment (Borovcanin et al., 2017), and is a mechanism that is being explored in experimental medicine studies (<https://gtr.ukri.org/projects?ref=MR%2FS037675%2F1>).

The IFN- γ cluster was comprised of patients with markedly elevated levels of IFN- γ compared with HC and all the other clusters, but it also had the lowest CRP levels of all groups and comparatively low levels of most other cytokines in our sample. Studies in patients with schizophrenia have shown that IFN- γ levels have an inverse relationship with whole-brain GMV and also left middle frontal gyrus thickness, potentially through alteration of the permeability of the blood brain barrier and activation of inflammatory T-cells and antigen-presenting cells along with the regulation of MHC class 2 genes (Abbott et al., 2006; Lesh et al., 2018). In contrast to previous studies, the patients with schizophrenia in our IFN- γ cluster showed the least GMV reduction compared with HCs and least neurocognitive deficits in language, attention, and verbal memory tasks. It also suggests that elevated IFN- γ may be beneficial in patients with schizophrenia, perhaps via helping to suppress viruses that can damage the brain (Dickerson et al., 2019). Altogether, our results point towards a potential preserved neuroanatomical and neurocognitive profile which is supported by recent studies reporting a positive effect of IFN- γ on symptoms (Vetlugina et al., 2016).

Furthermore, the fact that these patients still had hippocampal GMV loss suggests that this

loss is central to schizophrenia even in patients who are functionally more preserved, as proposed by others (Velakoulis et al., 1999). The fact that hippocampal GMV loss was present in this preserved group points to a modulatory hippocampal effect on the pathophysiology of schizophrenia (with patients showing more pronounced GMV alterations exhibiting a worse neurocognitive profile)(Grace, 2012; Van Rheenen et al., 2018). Similarly, previous evidence also suggests higher levels of IL-10, as seen in the Anti-Inflammatory cluster, may be protective against severe symptoms(Kay et al., 1987), also reflecting the relative sparing of GMV reduction seen in this cluster.

Patients in the Low Inflammation cluster revealed minimal inflammation compared with HCs and the rest of the clusters, which could be indicative of a less severe disease status or potentially successful treatment. Research has shown that the therapeutic efficacy of antipsychotic medication may be mediated by its effects on the basal ganglia with medicated patients exhibiting pallidal GMV increases compared with a placebo group(Chopra et al., 2021). Our GMV results show that patients in the Low Inflammation cluster exhibited GMV increases in the bilateral pallidum, consistent with a good treatment response. However, they still had temporal lobe volume loss.

Strengths and limitations

This study has several strengths including a large dataset, significance testing of the clustering solutions and exploration in an external dataset. Furthermore, advanced semi-supervised machine-learning (HYDRA) addresses challenges associated with conventional clustering methods by capturing multiple dimension of heterogeneity with varying numbers of hyperplanes. We demonstrated distinct brain structure patterns and neurocognitive performance (when the Low and Anti-Inflammatory clusters were combined) associated with subgroups. However results should be cautiously interpreted due to certain limitations. First,

the algorithm detected 8 clusters, three of these clusters had a very small number of patients, therefore rendering any further investigation or their structural or functional differences not possible. Those clusters may have been driven by unmeasured factors (e.g. infection) given that inflammatory levels were extremely high. Secondly, in the ASRB dataset, detailed positive symptom measures are not readily available, and thus the relation of inflammation to positive symptom severity is not captured. Finally, the BeneMin sample may not have had sufficient size and cytokine measurement sensitivity to identify all clusters observed in the ASRB dataset.

Conclusions

Identification of valid, reproducible subgroups of patients with schizophrenia based on inflammatory and neuroanatomical profile can aid in understanding putative and separable mechanisms of illness and pathogenic processes(Comer et al., 2020). This could contribute to the identification of subgroup candidates for targeted novel anti-inflammatory treatments in schizophrenia. Evidence for anti-inflammatory treatment in schizophrenia is mixed possibly due to the lack of immune subgroup identification for targeted treatment(Deakin et al., 2018). Our results show that it may be possible to identify reliable subgroups of schizophrenia patients based on accessible, peripheral measures of inflammation in large sample sizes using advanced clustering techniques. These results need to be validated in international datasets and explored in longitudinal designs to further establish their stability.

Declaration of Interests

Rachel Upthegrove reports grants from Medical Research Council, grants from National Institute for Health Research: Health Technology Assessment, grants from European Commission - Research: The Seventh Framework Programme, and personal speaker fees from Sunovion, outside the submitted work. Christos Pantelis has participated on Advisory Boards for Janssen-Cilag, Astra-Zeneca, Lundbeck, and Servier. He has received honoraria for talks presented at educational meetings organised by Astra-Zeneca, Janssen-Cilag, Eli-Lilly, Pfizer, Lundbeck and Shire. Nusrat Husain is a past Trustee of Manchester Global Foundation (MGF), The Pakistan Institute of Living and Learning (PILL), Abaseen Foundation UK and Lancashire Mind UK. He is executive member of the Academic Faculty of the Royal College of Psychiatrists, London. He is a NIHR Senior Investigator. In the last five years he has attended educational events organized by various pharmaceutical industries.

Data Sharing:

Requests for sharing the anonymized database should be addressed to the lead authors.

Acknowledgements:

We thank Roxanne Cadiz and Dr. Hayley North in for excellent research assistance and expertise in performing the cytokine and hsCRP assays.

HC		Cluster 1 (Low Inflammation)	Cluster 2 (CRP Cluster)	Cluster 3 (Classic Inflammation Cluster)	Cluster 4 (IFN- γ Cluster)	Cluster 5 (Anti- Inflammatory Cluster)	Cluster 6	Cluster 7	Cluster 8	$t/z/\chi^2/F$	P Value
Sample Sizes, No.	600	142	121	82	80	32	5	3	2		
Age, Mean (SD)	43.5 (13.1)	38.3 (11.1)	40.4 (10.7)	40.1 (10.8)	39.2 (11.4)	39.7 (9.5)	42.2 (7.9)	47 (4.2)	40.3 (16.5)	$F = .673$	.611
Sex (Male/Female)	240/360	105/37	67/54	57/25	49/31	19/13	3/2	1/1	2/1	$\chi^2 = 11.5$	.021
Freezer Time (Days, Mean (SD))	2427.9 (352.4)	2707.6 (412.3)	2497.0 (496.2)	2596.3 (464.6)	2632.6 (466.6)	2531.8 (383.8)	2763.2 (454.5)	2748.0 (426.9)	3039 (21.2)	$F = 3.8$	.005
Site No (Percentage)										$\chi^2 = 21.2$	.168
• Brisbane	• 193 (51.3)	• 44 (11.7)	• 49 (13.0)	• 42 (11.1)	• 34 (9.0)	• 11 (2.9)					
• Melbourne	• 178 (59.9)	• 40 (13.4)	• 28 (9.4)	• 19 (6.3)	• 14 (4.7)	• 11 (3.7)					
• Perth	• 97 (68.7)	• 17 (12.6)	• 12 (8.5)	• 6 (4.2)	• 7 (4.9)	• 2 (1.4)					
• Newcastle	• 55 (57.2)	• 16 (16.6)	• 11 (11.4)	• 1 (1.0)	• 8 (8.3)	• 5 (5.2)					
• Sydney	• 77 (49.0)	• 25 (15.9)	• 21 (13.3)	• 14 (8.9)	• 17 (10.8)	• 3 (1.9)					
CRP (pg/ml) (Mean (SD))	1.8 (2.3)	1.5 (1.3)	7.4 (3.0)*	2.6 (2.8)*	1.1 (1.2)	1.8 (1.2)	3.2 (3.9)	6.9 (1.5)	1.2 (0.4)		
IFN- γ (pg/ml) (Mean (SD))	29.7 (18.8)	18.4 (10.4)*	31.6 (15.3)	21.2 (12.3)	46.3 (16.9)*	36.3 (13.0)	19.3 (7.7)	15.3 (7.7)	22.4 (6.8)		
IL-10 (pg/ml) (Mean (SD))	34.3 (33.5)	26.3 (20.8)	34.5 (20.1)	31.6 (19.9)	34.2 (22.5)	89.0 (86.5)*	49.8 (24.3)	64.3 (87.2)	85.4 (36.2)		
IL-12 (pg/ml) (Mean (SD))	9.6 (6.3)	5.9 (3.3)*	10.3 (5.6)	7.2 (4.0)	14.9 (12.5)*	11.4 (3.5)	8.7 (3.0)	5.9 (2.1)	23.9 (1.2)		
IL-6 (pg/ml) (Mean (SD))	60.9 (289.4)	52.5 (185.7)	32.9 (86.6)	339.2 (547.7)*	22.7 (55.0)	40.7 (139.1)	2281.6 (79.6)	2250.6 (92.3)	1943.0 (510.4)		
IL-8 (pg/ml) (Mean (SD))	656.3 (1078.5)	529.8 (683.0)	534.0 (665.9)	3227.1 (1408.7)*	298.1 (501.2)	298.3 (355.4)	2472.9 (1601.8)	3258.0 (1008.2)	4306.0 (919.2)		
TNFe (pg/ml) (Mean (SD))	31.0 (183.6)	30.8 (55.9)	21.0 (21.3)	73.8 (100.4)	23.6 (42.2)	29.3 (49.7)	799.2 (308.4)	249.3 (77.3)	2131.8 (1038)		

Table 1. Sample Sociodemographics. Sample Sizes. Participants per Study Site. Age, Sex, Freezer Time. Cytokine concentrations (F values represent main effect of group on cytokine concentrations including HC, asterisks represent cytokines which are statistically significantly higher or lower to HC). Abbreviations: SD=Standard Deviation, pg=picograms, ml=millilitre.

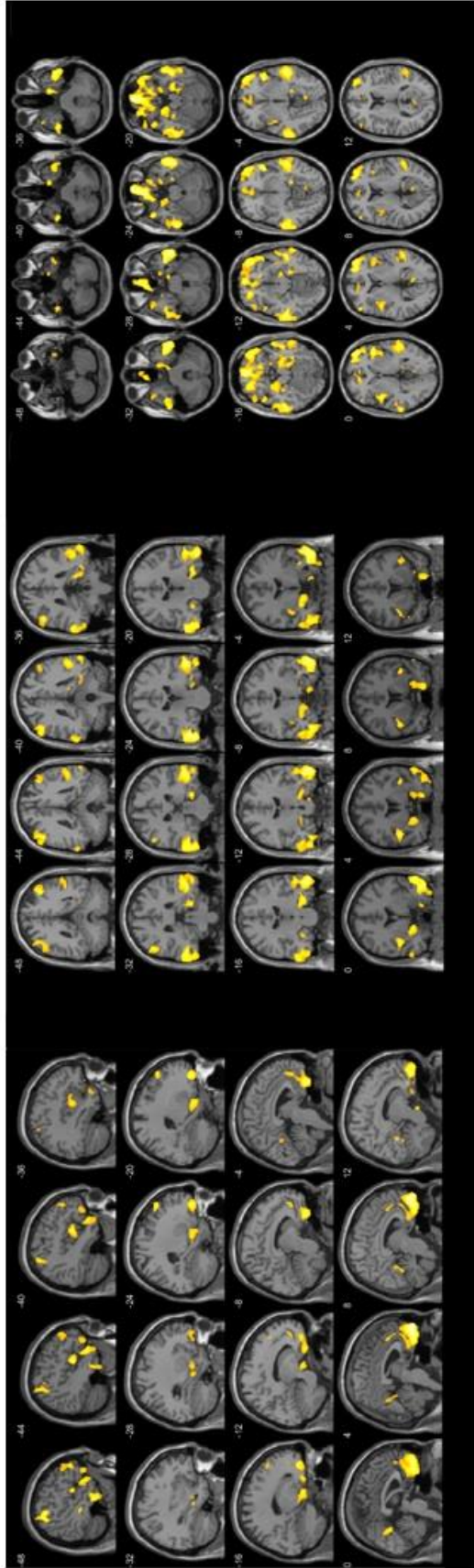


Figure 1. Classic Inflammation Cluster GMV Reductions Compared to HC. GMV reductions are observed in the Anterior Cingulate, Middle Frontal Gyrus, Inferior Frontal Gyrus, Superior Frontal Gyrus, Rectal Gyrus, Right Inferior frontal gyrus triangular part, Bilateral Insula, Middle Temporal Gyrus, Superior Temporal Gyrus, Inferior Temporal Gyrus, Bilateral Hippocampus, Middle Temporal Gyrus, Inferior Temporal Gyrus, Fusiform Gyrus, Superior Temporal Gyrus, Inferior Frontal Gyrus, Left Insula, Superior Frontal Gyrus, Inferior Parietal Lobule, and Right Inferior Parietal Lobe. Peak voxel MNI coordinates can be found in the supplement (table S1).

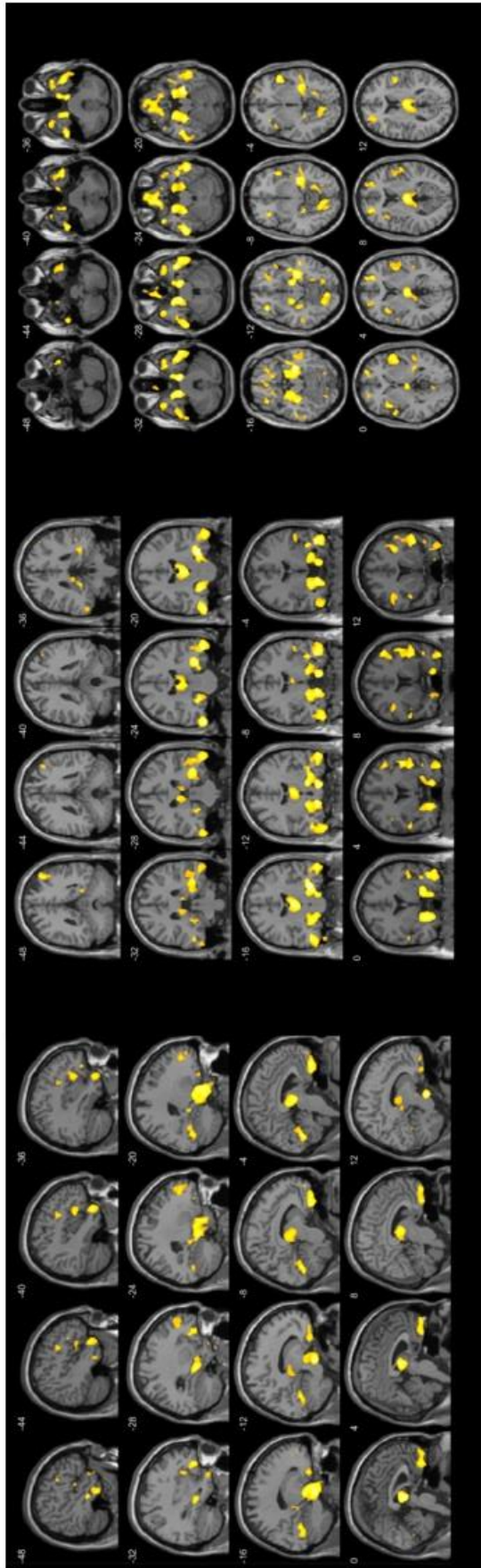


Figure 2. CRP Cluster GMV Reductions Compared to HC. GMV reductions are observed in the Parahippocampal Gyrus, Inferior Temporal Gyrus, Right Hippocampus, Middle Temporal Gyrus, Fusiform Gyrus, Amygdala, Bilateral Thalamus, Left Fusiform Gyrus, Inferior Frontal Gyrus, Right Insula, Left Lingual Gyrus, Superior Temporal Gyrus, Left Superior Frontal Gyrus, Left Insula, Inferior frontal gyrus opercular part Right, Right Superior Frontal Gyrus, Right Lingual Gyrus, Inferior Parietal Lobule, Angular Gyrus, Inferior frontal gyrus orbital part Left, and Inferior frontal gyrus opercular part Left. Peak voxel MNI coordinates can be found in the supplement (table S2).

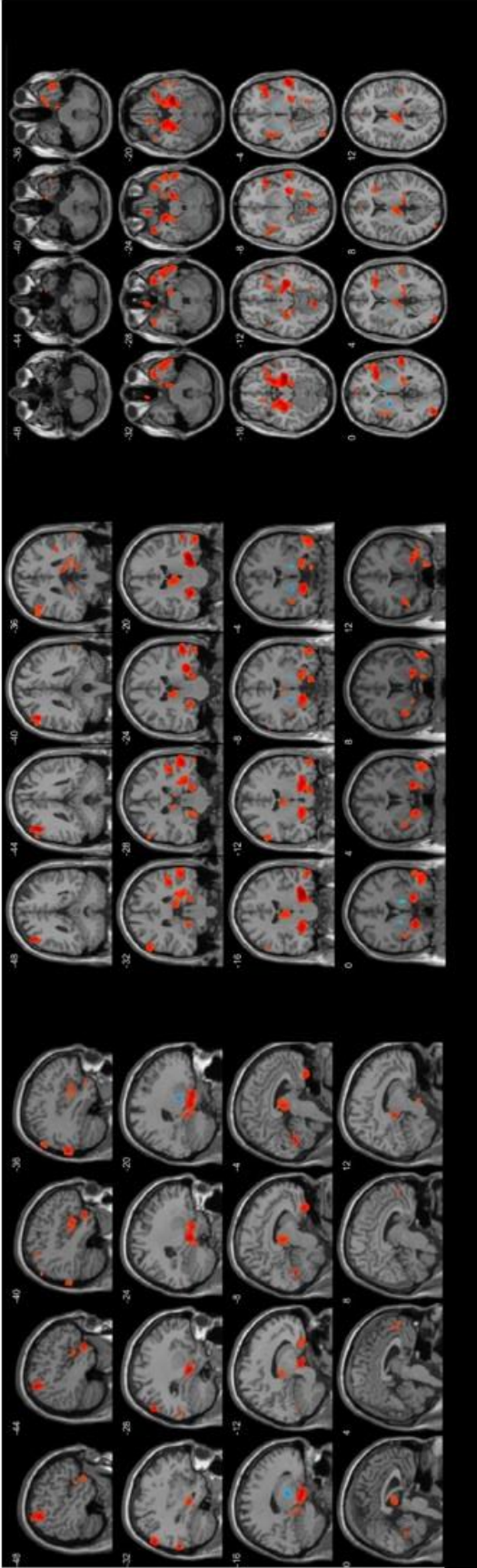


Figure 3. Low Inflammation Cluster GMV Reductions (red) and Increases (blue) Compared to HC. GMV reductions are observed in the

Parahippocampal Gyrus, Inferior Temporal Gyrus, Right Hippocampus, Middle Temporal Gyrus, Fusiform Gyrus, Right Amygdala, Inferior Temporal Gyrus Right, Left Hippocampus, Left Amygdala, Inferior Frontal Gyrus, Left Lingual Gyrus, Middle Occipital Gyrus, Right Superior Temporal Gyrus, Precuneus, and Inferior Parietal Lobule. GMV increases are observed in the bilateral pallidum. Peak voxel MNI coordinates can be found in the supplement (table S3).

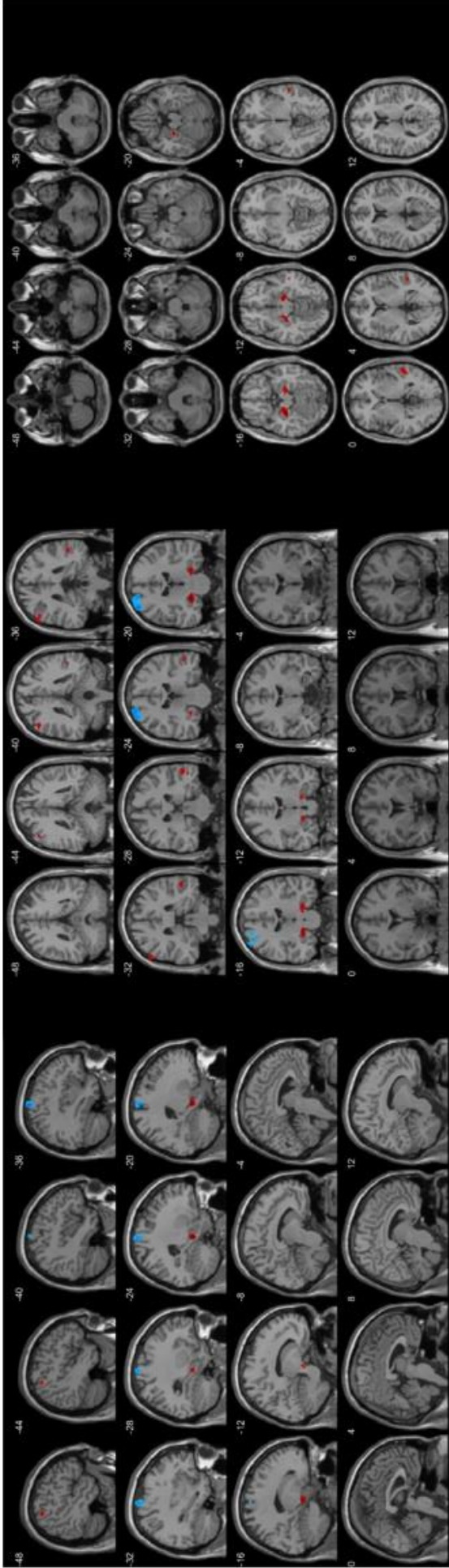


Figure 4. IFN- γ Cluster GMV Reductions (red) and Increases (blue) Compared to HC. GMV reductions are observed in the Parahippocampal Gyrus, Left Hippocampus, Parahippocampal Gyrus, Right Hippocampus, Middle Temporal Gyrus, Inferior Parietal Lobule GMV increases are observed in the precentral gyrus. Peak voxel MNI coordinates can be found in the supplement (table S4).

Chapter 7

Discussion

Summary of Chapter Findings

In chapter 1, I presented an overview of the core concepts explored in this thesis. First, this included an introduction to the history of schizophrenia broadening out to psychosis and of depression. I examined some of the current challenges to understanding of current neurobiological findings in the search for aetiology, with a focus on how neuroanatomical structure changes might reflect underlying aetiopathology. I present some evidence of the nosological, diagnostic, and treatment challenges that the presence of depression in the psychoses poses and discuss current models of understanding of mental disorders. After introducing explanations of the core concepts of machine learning, I then discuss its potential to help solve identified problems. In chapter 2, I outlined the aims and hypotheses of this PhD thesis. In Chapter 3, I performed a systematic review of structural neuroimaging studies that directly compared patients with depression and psychosis to identify structural neuroimaging similarities and differences. The systematic review results suggested potential transdiagnostic patterns of gray matter volume (GMV) and white matter volume (WMV) reductions in areas including the middle frontal gyrus (MFG), hippocampus, and posterior subgenual prefrontal cortex which could be further investigated as future biomarkers of comorbidity. However, these suggestions should be interpreted with caution since only seven studies fulfilled the systematic review criteria –which highlights the need for further research of the common and distinct brain features of psychosis and depression.

In Chapter 4, (Lalousis et al., 2021), I delineated the comorbidity between psychosis and depression using structural neuroimaging, phenotypic, and neurocognitive data with a series of supervised machine learning models. I thereby parsed the diagnostic heterogeneity within

and between psychotic and affective disorders in the early stages of illness using the multicentre, large PRONIA cohort. The results suggested that psychosis with comorbid depression exhibits disease processes that align more with non-comorbid (pure) depression rather than pure psychosis across the phenotypic, neurocognitive, and brain levels. These findings indicate a need to rethink current diagnostic classification to better reflect biological reality and potentially uncover phenotypes that may be hidden by current diagnostic structures. Eventually this could potentially lead to better informed and more targeted treatments. Motivated by the findings in Chapter 4, I undertook a bottom-up approach to the question of diagnostic classification of depression and psychosis, in Chapter 5. I used a novel semi-supervised machine learning algorithm, combining patients with depression and psychosis in one group. I aimed to identify cohesive clusters based on the patients' neuroanatomy thereby investigating the heterogeneity of brain structure across depression and psychosis. Two transdiagnostic clusters were identified, i.e., they both contained patients from both the depression and psychosis domains. One cluster exhibited an impaired neuroanatomical picture which was accompanied by negative symptomatology, poorer global and neurocognitive functioning, and increased levels of pro-inflammatory cytokines, while the other cluster showed a more preserved neuroanatomical picture and more depressive psychopathology. These clusters were tested for statistical significance and were externally validated using robust permutation and out-of-sample validation frameworks respectively. Importantly, we exhibited clinical utility as this chapter's results showed better prognostic performance of 9-month outcomes with a cluster distinction rather than a psychosis/depression distinction. Overall, Chapter 5's results challenge current diagnostic structures: as clusters identified in such a data-driven approach may, if further replicated, better reflect the biological reality of patients and help inform diagnostic structures which would be more informative for prognostic, and treatment choices. While these findings need

further replication, it appears that the identification of a transdiagnostic signature of poor prognosis is possible in early stages of mental disorders, thereby giving potential for new and targeted treatment strategies.

In Chapter 6, motivated by the previous chapter's findings of the association of markers of inflammation with an impaired neuroanatomical cluster, I further investigated the heterogeneity of inflammation in schizophrenia and its association to brain structure- starting this time with blood-based biomarker data. I implemented the same semi-supervised machine learning technique from Chapter 5 to identify clusters (subgroups) of people with schizophrenia that are similar in terms of inflammatory profiles -to the best of my knowledge this is the first time this advanced clustering has been applied to inflammation data in schizophrenia. Five main clusters were identified -each with a differing inflammatory and neuroanatomical profile. Specifically, three different types of clusters which had elevated pro-inflammatory markers and two clusters with low/elevated anti-inflammatory markers. All the clusters exhibited worse neurocognitive functioning compared with healthy controls. There was also a distinct between-cluster neurocognitive signature. Two of the clusters could also be detected in an external, younger dataset with patients in the clusters showing a similar neurocognitive pattern to the ones in the clusters of the discovery dataset. This chapter's results may reflect heterogeneity of inflammation pathways in a sub-group of people with schizophrenia and inflammation, currently masked in studies using more conventional techniques leading to less fine-grained separations such as low vs high inflammation. These findings have the potential to inform future targeted clinical trials and further dissect causal inflammatory pathways in subgroups of people with psychosis.

General Discussion

Depression and Psychosis

Schizophrenia is at the centre of a challenge to the validity of diagnostic categories in mental health disorders, given the increasing acknowledgement of a lack of clear clinical and neurobiological boundaries. A transdiagnostic approach, which suggests that some symptoms occur across diagnoses, is a potential solution to this challenge. Chapter 4's analysis in large recent onset psychosis and depression groups demonstrates that transdiagnostic clinical and neurocognitive features are accurately able to classify categorical diagnoses, in the absence of depression related co-morbidity. This suggests that there are identifiable 'pure' categorical groupings, however in early stages of illness co-morbidity is the norm. In the presence of co-morbidity there is an absence of hierarchy: a transdiagnostic signature is seen. However, these results highlight that both categorical and transdiagnostic approaches may ultimately fail, as neither recognise the possibility of pluripotent pathways that are both stage and context dependent. The implications are that personalised medicine approaches may be necessary, but with far more fine-grained characterisation of biological and psychological processes than is currently captured.

Fould's hierarchical theory, reflected in current nosological classification systems within psychiatry suggests that psychosis sits on the top of the hierarchy, with anxiety and depression symptoms of secondary importance (Foulds & Bedford, 1975). In Chapter 4, we found that rather than symptoms being unique to one diagnostic category, many are shared between two different categories and most importantly affect each one in a different way. This is convergent with the Research Domain Criteria (RDoC) project pillars (Cuthbert, 2014) Psychotic symptoms appeared to not alter the signature of ROD patients whilst depressive symptoms had a profound effect on ROP patients across the phenotypic,

neurocognitive, and brain levels. Such findings add evidence to the debate around the validity of the system on which the DSM is built upon (Stein et al., 2010) , suggesting that a descriptivist position in the diagnosis of mental disorder is not sufficient and that a novel multivariate look of mental disorder is more appropriate.

Comorbid symptoms have contributed to significant debates regarding heterogeneity in mental disorders, their function as prognostic indicators, and their place within hierarchical diagnostic structures. The clinical heterogeneity of mental disorders is indisputable, with no two patients presenting with the exact same constellation of symptoms, and symptoms change over time (Fusar-Poli et al., 2019). There may be significant interplay between sets or clusters of symptoms over the development of disorder from prodrome to onset to chronicity. For example, the frequency of depression within developing psychosis, and those in Clinical High Risk for psychosis, may rather than being ‘‘secondary’’ comorbid symptoms, be the primary driver. If identified correctly, rather than seen as comorbid and secondary, novel symptoms may be new targets that if treated, effectively ameliorate or drive the resolution of positive symptoms. Such observations give way to considerations that recent onset disorders may even constitute groups of generally common, phenotypically highly individual underlying aetiopathology, and those personalized treatments should be tailored accordingly.

In Chapter 4, the clinical SVM classification model found that different types of anhedonia (social and physical) were important in the classification of both psychosis and depression. Anhedonia has been suggested as a possible biomarker for depression (Der-Avakian & Markou, 2012; Hasler et al., 2004) and has been found to be associated with decreased activation in in ventral basal ganglia areas, the dorsal anterior cingulate, middle frontal gyrus, and medial frontal gyrus both in schizophrenia and depression (B. Zhang et al., 2016). Chapter 4’s neuroimaging and clinical models revealed that orbitofrontal areas and physical

anhedonia were weighted in the classification to the pure ROD group suggesting a potential association between physical anhedonia and lower orbitofrontal GMV in depression.

As we move away from traditional descriptivist approaches to diagnosis there may not be a simple transdiagnostic solution even at the symptom level. Whether categorical schizophrenia is eventually defined by the ‘p’ factor or dimensional symptom approaches, I suggest that there are unseen clinical, potentially relevant features that lie within a mechanistically transdiagnostic picture.

Heterogeneity and New Diagnostic Approaches

The Kraepelinian dichotomy has long been challenged and studies have shown that our understanding of the clinical and neurobiological distinction of depression and psychosis may be particularly challenging during early phases of illness (Birchwood et al., 2005; Craddock & Owen, 2010; Koutsouleris et al., 2015; Upthegrove et al., 2017). The usefulness and aetiopathogenetic basis of the current diagnostic system is questionable yet these diagnoses have remained the same for around 100 years due to the lack of accepted (neuro)biological or genetic markers and the lack of better alternatives (Cuthbert, 2014; First et al., 2018; Fusar-Poli, et al., 2019; Linden, 2012). This lack of biological validity of diagnostic groups is thought to be one of the major reasons for poor biomedical translation resulting in better outcomes for patients with mental illness (Brückl et al., 2020; Kapur et al., 2012; Kendler, 2013). Indeed, only 20% of people with psychosis and 25% of people with depression achieve full remission and response to pharmacological treatment, with the remainder achieving partial response or response without remission (S. Brown et al., 2010; Health (UK), 2014; Kern et al., 2020; Owen et al., 2016). Identifying biologically driven illness models, able to relate to those at highest risk of chronicity would allow new and targeted treatments to be delivered early. Targeted early intervention can prevent poor outcomes in patients with

recent onset depression and psychosis (McGorry et al., 2008; Murphy & Brewer, 2011). However, identifying those patients who are on a path to chronic disability is still difficult in both psychosis and depression. In depression, predicting illness course at the individual level is moderate at best and in psychosis it has been suggested that there is a need for neurobiological and psychosocial relevance to be integrated in predictive modelling, harnessing large sample sizes and long observational periods (Dinga et al., 2018; Klosterkötter et al., 2011).

The Kraepelinian dichotomy between dementia praecox and "manic-depressive insanity" is a distinction between severe affective vs non-affective mental illness. Dementia praecox does not directly map to non-affective psychosis, however the Verrücktheit (chronic non-affective psychoses) made distinct in Kraepelin's first Edition (1883) lead to the (mis) understanding that schizophrenia was non-affective. In Chapter 5 I extended the challenge of this dichotomy into early stages of affective and non-affective recent onset disorders. The tension of depression in schizophrenia presented a taxonomic challenge that was perpetuated with "true schizophrenia" being considered as non-affective. Carpenter and Fisher outline that, as defined by Kraepelin's 6th edition core features -negative, hebephrenia and catatonia- schizophrenia persists and Chapter 5's results would also not refute this. Although there were no measures of catatonia, the cluster labelled as "impaired" has more cognitive symptoms and a brain signature identified in the chronic replication sample -of patients with depression and schizophrenia.

There is a body of literature, which challenges a dichotomy between affective and non-affective psychoses whilst still maintaining potential distinction between deteriorating or primary established cases. Craddock and Owen's position in 2010 was that "Although a great deal of work remains to be undertaken in delineating the pathogenically relevant DNA variants and the biological mechanisms by which they influence disease risk, the studies

described above indicate that schizophrenia and bipolar disorder (and recurrent depression) do not ‘breed true’, but have an overlap in genetic risk and are therefore likely to share some aspects of pathogenesis.” (Craddock & Owen, 2010). Since this time studies of genomic risk across disorders have increasingly demonstrated transdiagnostic genetic risk for psychiatric illness (Verduijn et al., 2017; Wray et al., 2018). Depression has been shown to be significant in the prediction of transition to FEP, poor clinical outcomes, quality of life, and suicide in schizophrenia (Upthegrove et al., 2017). Depressive symptoms are highly interrelated with both positive symptoms and negative symptoms of motivation and interest, and might therefore be best conceptualised as intrinsic to early phases of schizophrenia spectrum disorders (Herniman et al., 2021). Dwyer et al., (2020) previously suggested a need to consider short-term to medium-term presentations and focus on functioning in psychosis patients stratified in the depressive and severe psychosis subgroups identified with data-driven clustering using high-dimensional clinical data. The Kraepelinian dichotomy debate is not one whose outcome has been decisively concluded, but Chapter 5 results add to this narrative by showcasing that in early onset there is a cluster characterised by poor outcome across the depressive and psychotic phenotypes. Fischer and Carpenter correctly stated in 2009 that data to categorically falsify the Kraepelinian dichotomy had not yet been produced but they also make a call for heterogeneity reduction in order for the concept to be successfully investigated. I believe the paper presented in Chapter 5 adds to this evidence.

Biological Relevance and Precision Psychiatry

Current nosological categories are not reflective of underlying brain pathology. In Chapter 5, we identify a transdiagnostic cluster with an impaired phenotype in recent onset patients with psychosis and depression, that had some distinct inflammatory characteristics. Psychiatric illnesses are not a single variable problem and I aimed to further elucidate whether phenotypic and cognitive data adds to the biologically based clusters in chapters 4 and 5

using a multimodal data approach. While these clusters are initially derived from a single variable class (inflammation), they are meaningful because of the distinct additional data signatures. As clusters have been derived in recent onset samples using structural MRI in Chapter 5 and cytokine data in Chapter 6, this may also add to the evidence of the utility of MRI and blood sampling in potentially creating tools to eventually help clinician prediction and decision making for targeted treatment choices. Reducing heterogeneity and creating diagnostic structures with increased biological relevance is important for biomedical translation as we move towards precision psychiatry. Previous research which reduced heterogeneity in schizophrenia has identified that affective symptoms in schizophrenia are related to the hippocampus and the putamen/pallidum (Liu et al., 2021). Preserved basal ganglia structure has also been associated with a potentially non-progressive course (Chand et al., 2020). In Chapter 5 reducing neuroanatomical heterogeneity resulted in diagnostic clusters with increased biological relevance and prognostic potential. The preserved cluster showed less GMV loss and was associated with a less severe clinical picture and more depressive symptoms. It appears that heterogeneity is driven by different aetiopathological pathways which are reflected in brain structure and relate to different symptom profiles. Interestingly, the impaired cluster had associated pro-inflammatory elevation and for both clusters when adding cytokine data to the prognostic models, predictive accuracy increased. The relationship between inflammation and brain structure can give us an insight into potential mechanisms for heterogeneous neuroanatomical pathways. Elevated pro-inflammatory cytokines which have responded to genomic predisposition and/or response to environmental factors may result in astrocytic dysfunction and/or microglial activation. The subsequent dendritic pruning and synaptic changes then lead to GMV loss (Corsi-Zuelli & Deakin, 2021). However, heterogeneity is inherent in inflammatory pathways to psychiatric illness.

Longitudinal research has shown potential causal mechanisms between increased serum concentration of IL-6 and CRP in childhood and adolescence and development of schizophrenia and related psychoses (Khandaker et al., 2014; Wium-Andersen et al., 2014). Elevated levels of IL-6, TNF- α , IL-1 β , and IFN- γ and reduced levels of IL-10 have been associated with antipsychotic-naïve first-episode psychosis (Upthegrove et al., 2014) and acute psychotic relapse (B. J. Miller et al., 2011). Poor treatment response has been linked to elevated levels of specific biomarkers such as IL-6 and IFN- γ (Mondelli et al., 2015). Reducing heterogeneity in the field of inflammation is necessary if the promise of novel schizophrenia treatments is to be realized. In Chapter 6 I showed that the subdivision of inflammatory responses in high vs low is missing out on a heterogeneity that is present in both high and low inflammation groups. Reducing heterogeneity in inflammatory pathways in schizophrenia can help us understand the different ways in which inflammation can lead to heterogeneous GMV loss with associated heterogeneous symptom profiles. Furthermore, the elucidation of different inflammatory pathways to schizophrenia and their association to brain structure has the potential to inform future clinical trials and drug development for individualized treatment.

Clinical Applicability

The results of this thesis challenge the current nosological status quo in psychiatry and show promise of clinical applicability. While the machine learning models developed here cannot be directly applied in clinical settings and would first need extensive rigorous testing, they provide a solid basis for the translation of neuroscientific findings into practice. In Chapter 5, attempting to simulate clinical setting decision-making based on the current nosological system, I compared the accuracy of 9-month symptomatological prediction between traditional diagnostic categories and neurobiologically-based ones. I showed that the

neurobiologically-based diagnostic categories outperform the traditional ones. Such models need to be prospectively validated in readily available and routinely collected public healthcare datasets to examine their utility in the real world. The advantage of machine learning methods being able to make single patient predictions mean that small scale clinical trials would be adequate to showcase the utility of such models.

In Chapter 4, I showed that, potentially, the frequency of depression within developing psychosis may be the primary driver of illness rather than secondary comorbid symptomatology. Our models using transdiagnostic symptom data exhibited the possibility of pluripotent pathways to illness that are both stage and context dependent. These transdiagnostic symptoms can be targeted in early mental illness or even be new treatment targets. Indeed, using network modelling, the relationships between transdiagnostic symptoms and patients' outcomes has been shown to potentially lead to global symptom improvements and better recovery (Griffiths et al., 2021).

In Chapter 6, I exhibited that machine learning models can be used to elucidate different inflammatory subgroups in schizophrenia. The field of inflammation in psychiatric illness offers exciting avenues for new or repurposed treatments. Machine learning models developed in Chapter 6 can be used to stratify clinical trials for the targeted treatment of anti-inflammatory medication and eventually be used in clinical practice to aid novel treatment decision-making. Ultimately, after rigorous testing, the findings of this thesis can help clinical practice steer away from the trial and error “one-size-fits-all” approach and move to truly precision medicine.

Strengths and Limitations

This PhD thesis contains several strengths such as the use of international multi-site datasets such as PRONIA, MUC, MCIC, COBRE, ASRB, and BeneMin. I also did not follow a dogmatic approach in the types of machine learning used as evidenced by the use of both supervised and semi-supervised machine learning in a multitude of programming languages and packages such as MATLAB, R, Python, Powershell, NeuroMiner, HYDRA, ComBat, and SigClust, as well as the coding of my own permutation frameworks and machine learning pipelines. A significant strength of this PhD thesis is that the results were related to clinical practice and attempted to add to nosological and aetiological theory. Nevertheless, its results need to be interpreted with caution due to certain limitations. Clustering methods are inherently limited by the data used to develop the models. For example, neuroimaging data might produce very different clusters with different associations with symptom and cognitive data than inflammatory marker input data. This leads to the question of how can divergent input data be consolidated into clinically useful and meaningful models. I addressed this limitation by always relating initial findings to several levels (brain, phenotypic, neurocognitive, inflammatory). Consensus clustering is a method in which different clustering method results get combined to produce a single consensus model. This could be used in future studies where multimodal data can be leveraged before combined into a consensus model -this has already been successfully implemented in the cancer field (The Cancer Genome Atlas Research Network, 2015). Furthermore, the inference of underlying pathophysiological processes of the study population in this PhD thesis has been based on cross-sectional data. To confidently infer underlying pathophysiology of disease processes longitudinal examination is needed. In chapters 4 and 5 where longitudinal data were available, I examined longitudinal trajectories to further strengthen the inferences made in this PhD thesis. Finally, it is important that social and environmental factors such as

urbanicity, pervasive disadvantage, marginalization and adversity, which have been shown to significantly contribute to mental health disorders, are included in machine learning models that attempt to improve the current diagnostic and prognostic tools in psychiatry. Throughout this PhD thesis, whenever machine learning models have been developed using neurobiologically-based data, there has been an attempt to relate findings to multiple levels that span the phenotypic, neurocognitive, and environmental domains when the data allowed for it. This is important for the future translation of machine learning prediction as it encompasses the complexity of illness outcomes and presentations.

Conclusions

Heterogeneity and comorbidity within and between psychosis and depression has led to the creation of new diagnostic categories and poor biomedical translation of neuroscientific studies. In this PhD thesis, machine learning was utilised to elucidate this heterogeneity, reduce it, and identify biologically relevant and transdiagnostic subtypes that could become potential candidates for targeted treatment. The results are promising and challenge the current nosological system. In 2017, Thomas Insel suggested that the biomedical framework that he adopted while directing the NIMH failed to “move the needle” in improving the lives of people with mental illnesses. If precision (and impactful) medicine is to be achieved in psychiatry, neurobiological data from large sample sizes need to be submitted to meticulous and novel methodological approaches and related to all levels, and the results need to ambitiously challenge current dogma.

References

- Abbott, N. J., Rönnbäck, L., & Hansson, E. (2006). Astrocyte-endothelial interactions at the blood-brain barrier. *Nature Reviews. Neuroscience*, 7(1), 41–53.
<https://doi.org/10.1038/nrn1824>
- Alvarez, E., Garcia-Ribera, C., Torrens, M., Udina, C., Guillamat, R., & Casas, M. (1987). Premorbid adjustment scale as a prognostic predictor for schizophrenia. *The British Journal of Psychiatry: The Journal of Mental Science*, 150, 411.
<https://doi.org/10.1192/bjp.150.3.411>
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association.
- American Psychiatric Association (Ed.). (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV ; includes ICD-9-CM codes effective 1. Oct. 96* (4. ed., 7. print).
- American Psychiatric Association & American Psychiatric Association (Eds.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Association.
- Andreasen, N. C. (1989). The Scale for the Assessment of Negative Symptoms (SANS): Conceptual and theoretical foundations. *The British Journal of Psychiatry. Supplement*, 7, 49–58.
- Arnold, S. J. M., Ivleva, E. I., Gopal, T. A., Reddy, A. P., Jeon-Slaughter, H., Sacco, C. B., Francis, A. N., Tandon, N., Bidesi, A. S., Witte, B., Poudyal, G., Pearlson, G. D., Sweeney, J. A., Clementz, B. A., Keshavan, M. S., & Tamminga, C. A. (2015). Hippocampal Volume Is Reduced in Schizophrenia and Schizoaffective Disorder But Not in Psychotic Bipolar I Disorder Demonstrated by Both Manual Tracing and Automated Parcellation (FreeSurfer). *Schizophrenia Bulletin*, 41(1), 233–249.
<https://doi.org/10.1093/schbul/sbu009>

- Arnone, D., McIntosh, A. M., Ebmeier, K. P., Munafò, M. R., & Anderson, I. M. (2012). Magnetic resonance imaging studies in unipolar depression: Systematic review and meta-regression analyses. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, 22(1), 1–16.
<https://doi.org/10.1016/j.euroneuro.2011.05.003>
- Baca-Garcia, E., Perez-Rodriguez, M. M., Basurte-Villamor, I., Fernandez del Moral, A. L., Jimenez-Arriero, M. A., Gonzalez de Rivera, J. L., Saiz-Ruiz, J., & Oquendo, M. A. (2007). Diagnostic stability of psychiatric disorders in clinical practice. *The British Journal of Psychiatry: The Journal of Mental Science*, 190, 210–216.
<https://doi.org/10.1192/bjp.bp.106.024026>
- Ballard, E. D., Henter, I. D., & Zarate, C. A. (2019). Chapter 1 - The Classification of Depression: Embracing Phenotypic Heterogeneity in the Era of the RDoC. In J. Quevedo, A. F. Carvalho, & C. A. Zarate (Eds.), *Neurobiology of Depression* (pp. 1–8). Academic Press. <https://doi.org/10.1016/B978-0-12-813333-0.00001-9>
- Barkus, E., & Badcock, J. C. (2019). A Transdiagnostic Perspective on Social Anhedonia. *Frontiers in Psychiatry*, 10, 216. <https://doi.org/10.3389/fpsyt.2019.00216>
- Beck, A., Steer, R., & Brown, G. (1996). *Manual for the Beck Depression Inventory-II (BDI-II)*. <https://www.scienceopen.com/document?vid=9feb932d-1f91-4ff9-9d27-da3bda716129>
- Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996). Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *Journal of Personality Assessment*, 67(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13
- Becker, H. E., Nieman, D. H., Dingemans, P. M., van de Fliert, J. R., De Haan, L., & Linszen, D. H. (2010). Verbal fluency as a possible predictor for psychosis. *European*

- Psychiatry: The Journal of the Association of European Psychiatrists*, 25(2), 105–110. <https://doi.org/10.1016/j.eurpsy.2009.08.003>
- Beer, M. D. (1995). Psychosis: From mental disorder to disease concept. *History of Psychiatry*, 6(22), 177–200. <https://doi.org/10.1177/0957154X9500602204>
- Bernstein, D. P., Fink, L., Handelsman, L., Foote, J., Lovejoy, M., Wenzel, K., Sapareto, E., & Ruggiero, J. (1994). Initial reliability and validity of a new retrospective measure of child abuse and neglect. *The American Journal of Psychiatry*, 151(8), 1132–1136. <https://doi.org/10.1176/ajp.151.8.1132>
- Bero, L. (2006). Evaluating Systematic Reviews and Meta-Analyses. *Journal of Law and Policy*, 14(2). <https://brooklynworks.brooklaw.edu/jlp/vol14/iss2/4>
- Birchwood, M., Iqbal, Z., & Upthegrove, R. (2005). Psychological pathways to depression in schizophrenia: Studies in acute psychosis, post psychotic depression and auditory hallucinations. *European Archives of Psychiatry and Clinical Neuroscience*, 255(3), 202–212. <https://doi.org/10.1007/s00406-005-0588-4>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Boerrigter, D., Weickert, T. W., Lenroot, R., O'Donnell, M., Galletly, C., Liu, D., Burgess, M., Cadiz, R., Jacomb, I., Catts, V. S., Fillman, S. G., & Weickert, C. S. (2017). Using blood cytokine measures to define high inflammatory biotype of schizophrenia and schizoaffective disorder. *Journal of Neuroinflammation*, 14(1), 188. <https://doi.org/10.1186/s12974-017-0962-y>
- Borovcanin, M. M., Jovanovic, I., Radosavljevic, G., Pantic, J., Minic Janicijevic, S., Arsenijevic, N., & Lukic, M. L. (2017). Interleukin-6 in Schizophrenia—Is There a Therapeutic Relevance? *Frontiers in Psychiatry*, 8, 221. <https://doi.org/10.3389/fpsy.2017.00221>

- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13.
<https://doi.org/10.1002/wps.20375>
- Brand, S. J., Möller, M., & Harvey, B. H. (2015). A Review of Biomarkers in Mood and Psychotic Disorders: A Dissection of Clinical vs. Preclinical Correlates. *Current Neuropsychopharmacology*, 13(3), 324–368.
<https://doi.org/10.2174/1570159X13666150307004545>
- Broome, M. R., He, Z., Iftikhar, M., Eyden, J., & Marwaha, S. (2015). Neurobiological and behavioural studies of affective instability in clinical populations: A systematic review. *Neuroscience and Biobehavioral Reviews*, 51, 243–254.
<https://doi.org/10.1016/j.neubiorev.2015.01.021>
- Brown, A. M., Conn, I., Boerrigter, D., Shannon Weickert, C., & Purves-Tyson, T. D. (2022). Maternal immune activation with high molecular weight poly(I:C) in Wistar rats leads to elevated immune cell chemoattractants. *Journal of Neuroimmunology*, 364, 577813. <https://doi.org/10.1016/j.jneuroim.2022.577813>
- Brown, S., Kim, M., Mitchell, C., & Inskip, H. (2010). Twenty-five year mortality of a community cohort with schizophrenia. *The British Journal of Psychiatry: The Journal of Mental Science*, 196(2), 116–121. <https://doi.org/10.1192/bjp.bp.109.067512>
- Brückl, T. M., Spoormaker, V. I., Sämann, P. G., Brem, A.-K., Henco, L., Czamara, D., Elbau, I., Grandi, N. C., Jollans, L., Kühnel, A., Leuchs, L., Pöhlchen, D., Schneider, M., Tontsch, A., Keck, M. E., Schilbach, L., Czisch, M., Lucae, S., Erhardt, A., & Binder, E. B. (2020). The biological classification of mental disorders (BeCOME) study: A protocol for an observational deep-phenotyping study for the identification of biological subtypes. *BMC Psychiatry*, 20(1), 213. <https://doi.org/10.1186/s12888-020-02541-z>

- Buka, S. L., Cannon, T. D., Torrey, E. F., & Yolken, R. H. (2008). Maternal Exposure to Herpes Simplex Virus and Risk of Psychosis Among Adult Offspring. *Biological Psychiatry*, 63(8), 809–815. <https://doi.org/10.1016/j.biopsych.2007.09.022>
- Bulzacka, E., Boyer, L., Schürhoff, F., Godin, O., Berna, F., Brunel, L., Andrianarisoa, M., Aouizerate, B., Capdevielle, D., Chéreau-Boudet, I., Chesnoy-Servanin, G., Danion, J.-M., Dubertret, C., Dubreucq, J., Faget, C., Gabayet, F., Le Gloahec, T., Llorca, P.-M., Mallet, J., ... Fond, G. (2016). Chronic Peripheral Inflammation is Associated With Cognitive Impairment in Schizophrenia: Results From the Multicentric FACE-SZ Dataset. *Schizophrenia Bulletin*, 42(5), 1290–1302. <https://doi.org/10.1093/schbul/sbw029>
- Bürgy, M. (2008). The Concept of Psychosis: Historical and Phenomenological Aspects. *Schizophrenia Bulletin*, 34(6), 1200–1210. <https://doi.org/10.1093/schbul/sbm136>
- Busatto, G. F. (2013). Structural and functional neuroimaging studies in major depressive disorder with psychotic features: A critical review. *Schizophrenia Bulletin*, 39(4), 776–786. <https://doi.org/10.1093/schbul/sbt054>
- Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- Campbell, S., & MacQueen, G. (2004). The role of the hippocampus in the pathophysiology of major depression. *Journal of Psychiatry and Neuroscience*, 29(6), 417–426.
- Cannon-Spoor, H. E., Potkin, S. G., & Wyatt, R. J. (1982). Measurement of Premorbid Adjustment in Chronic Schizophrenia. *Schizophrenia Bulletin*, 8(3), 470–484. <https://doi.org/10.1093/schbul/8.3.470>

- Carpenter, W. T., Heinrichs, D. W., & Wagman, A. M. (1988). Deficit and nondeficit forms of schizophrenia: The concept. *The American Journal of Psychiatry*, 145(5), 578–583.
<https://doi.org/10.1176/ajp.145.5.578>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, 2(2), 119–137.
<https://doi.org/10.1177/2167702613497473>
- Chand, G. B., Dwyer, D. B., Erus, G., Sotiras, A., Varol, E., Srinivasan, D., Doshi, J., Pomponio, R., Pigoni, A., Dazzan, P., Kahn, R. S., Schnack, H. G., Zanetti, M. V., Meisenzahl, E., Busatto, G. F., Crespo-Facorro, B., Pantelis, C., Wood, S. J., Zhuo, C., ... Davatzikos, C. (2020). Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain*, 143(3), 1027–1038.
<https://doi.org/10.1093/brain/awaa025>
- Chapelle, O., Scholkopf, B., & Zien, A. (Eds.). (2010). *Semi-Supervised Learning* (1st edition). The MIT Press.
- Charlson, F. J., Ferrari, A. J., Santomauro, D. F., Diminic, S., Stockings, E., Scott, J. G., McGrath, J. J., & Whiteford, H. A. (2018). Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. *Schizophrenia Bulletin*, 44(6), 1195–1203. <https://doi.org/10.1093/schbul/sby058>
- Chen, A., & Zhu, W. (2001). Revisiting the Assumptions for Inferential Statistical Analyses: A Conceptual Guide. *Quest*, 53(4), 418–439.
<https://doi.org/10.1080/00336297.2001.10491756>
- Chen, G., Guo, Y., Zhu, H., Kuang, W., Bi, F., Ai, H., Gu, Z., Huang, X., Lui, S., & Gong, Q. (2017). Intrinsic disruption of white matter microarchitecture in first-episode, drug-

- naive major depressive disorder: A voxel-based meta-analysis of diffusion tensor imaging. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 76, 179–187. <https://doi.org/10.1016/j.pnpbp.2017.03.011>
- Chong, H. Y., Teoh, S. L., Wu, D. B.-C., Kotirum, S., Chiou, C.-F., & Chaiyakunapruk, N. (2016). Global economic burden of schizophrenia: A systematic review. *Neuropsychiatric Disease and Treatment*, 12, 357–373. <https://doi.org/10.2147/NDT.S96649>
- Chopra, S., Fornito, A., Francey, S. M., O'Donoghue, B., Cropley, V., Nelson, B., Graham, J., Baldwin, L., Tahtalian, S., Yuen, H. P., Allott, K., Alvarez-Jimenez, M., Harrigan, S., Sabaroedin, K., Pantelis, C., Wood, S. J., & McGorry, P. (2021). Differentiating the effect of antipsychotic medication and illness on brain volume reductions in first-episode psychosis: A Longitudinal, Randomised, Triple-blind, Placebo-controlled MRI Study. *Neuropsychopharmacology*, 46(8), 1494–1501. <https://doi.org/10.1038/s41386-021-00980-0>
- Chuang, J.-Y., Murray, G. K., Metastasio, A., Segarra, N., Tait, R., Spencer, J., Ziauddeen, H., Dudas, R. B., Fletcher, P. C., & Suckling, J. (2014). Brain structural signatures of negative symptoms in depression and schizophrenia. *Frontiers in Psychiatry*, 5, 116. <https://doi.org/10.3389/fpsy.2014.00116>
- Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, 18(2), 72–145. <https://doi.org/10.1177/1529100617727266>
- Cole, J. D., & Kazarian, S. S. (1988). The Level of Expressed Emotion Scale: A new measure of expressed emotion. *Journal of Clinical Psychology*, 44(3), 392–397.

[https://doi.org/10.1002/1097-4679\(198805\)44:3<392::AID-JCLP2270440313>3.0.CO;2-3](https://doi.org/10.1002/1097-4679(198805)44:3<392::AID-JCLP2270440313>3.0.CO;2-3)

Collins, J. A., & Fauser, B. C. J. M. (2005). Balancing the strengths of systematic and narrative reviews. *Human Reproduction Update*, 11(2), 103–104.

<https://doi.org/10.1093/humupd/dmh058>

Comer, A. L., Carrier, M., Tremblay, M.-È., & Cruz-Martín, A. (2020). The Inflamed Brain in Schizophrenia: The Convergence of Genetic and Environmental Risk Factors That Lead to Uncontrolled Neuroinflammation. *Frontiers in Cellular Neuroscience*, 14, 274. <https://doi.org/10.3389/fncel.2020.00274>

Conley, R. R., Ascher-Svanum, H., Zhu, B., Faries, D., & Kinon, B. J. (2007). The Burden of Depressive Symptoms in the Long-Term Treatment of Patients With Schizophrenia. *Schizophrenia Research*, 90(1–3), 186–197.

<https://doi.org/10.1016/j.schres.2006.09.027>

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Foa, E., & Weisler, R. H. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. *The British Journal of Psychiatry: The Journal of Mental Science*, 176, 379–386. <https://doi.org/10.1192/bjp.176.4.379>

Cornblatt, B. A., Auther, A. M., Niendam, T., Smith, C. W., Zinberg, J., Bearden, C. E., & Cannon, T. D. (2007). Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophrenia Bulletin*, 33(3), 688–702. <https://doi.org/10.1093/schbul/sbm029>

Cornblatt, B. A., Risch, N. J., Faris, G., Friedman, D., & Erlenmeyer-Kimling, L. (1988). The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families. *Psychiatry Research*, 26(2), 223–238.

[https://doi.org/10.1016/0165-1781\(88\)90076-5](https://doi.org/10.1016/0165-1781(88)90076-5)

- Correll, C. U., & Schooler, N. R. (2020). Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. *Neuropsychiatric Disease and Treatment*, 16, 519–534. <https://doi.org/10.2147/NDT.S225643>
- Corsi-Zuelli, F., & Deakin, B. (2021). Impaired regulatory T cell control of astroglial overdrive and microglial pruning in schizophrenia. *Neuroscience & Biobehavioral Reviews*, 125, 637–653. <https://doi.org/10.1016/j.neubiorev.2021.03.004>
- Coryell, W., Nopoulos, P., Drevets, W., Wilson, T., & Andreasen, N. C. (2005). Subgenual prefrontal cortex volumes in major depressive disorder and schizophrenia: Diagnostic specificity and prognostic implications. *The American Journal of Psychiatry*, 162(9), 1706–1712. <https://doi.org/10.1176/appi.ajp.162.9.1706>
- Costa, P. T., McCrae, R. R., & Psychological Assessment Resources, I. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources.
- Craddock, N., & Owen, M. J. (2010). The Kraepelinian dichotomy – going, going... But still not gone. *The British Journal of Psychiatry*, 196(2), 92–95. <https://doi.org/10.1192/bjp.bp.109.073429>
- Cui, Y., Dong, J., Yang, Y., Yu, H., Li, W., Liu, Y., Si, J., Xie, S., Sui, J., Lv, L., & Jiang, T. (2020). White matter microstructural differences across major depressive disorder, bipolar disorder and schizophrenia: A tract-based spatial statistics study. *Journal of Affective Disorders*, 260, 281–286. <https://doi.org/10.1016/j.jad.2019.09.029>
- Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1), 28–35. <https://doi.org/10.1002/wps.20087>
- Davison, K. (2006). Historical aspects of mood disorders. *Psychiatry*, 5(4), 115–118. <https://doi.org/10.1383/psyt.2006.5.4.115>

- de Boer, N. S., de Bruin, L. C., Geurts, J. J. G., & Glas, G. (2021). The Network Theory of Psychiatric Disorders: A Critical Assessment of the Inclusion of Environmental Factors. *Frontiers in Psychology*, 12.
<https://www.frontiersin.org/article/10.3389/fpsyg.2021.623970>
- Deakin, B., Suckling, J., Barnes, T. R. E., Byrne, K., Chaudhry, I. B., Dazzan, P., Drake, R. J., Giordano, A., Husain, N., Jones, P. B., Joyce, E., Knox, E., Krynicki, C., Lawrie, S. M., Lewis, S., Lisiecka-Ford, D. M., Nikkheslat, N., Pariante, C. M., Smallman, R., ... Dunn, G. (2018). The benefit of minocycline on negative symptoms of schizophrenia in patients with recent-onset psychosis (BeneMin): A randomised, double-blind, placebo-controlled trial. *The Lancet Psychiatry*, 5(11), 885–894.
[https://doi.org/10.1016/S2215-0366\(18\)30345-6](https://doi.org/10.1016/S2215-0366(18)30345-6)
- Dean, J., & Keshavan, M. (2017). The neurobiology of depression: An integrated view. *Asian Journal of Psychiatry*, 27, 101–111. <https://doi.org/10.1016/j.ajp.2017.01.025>
- Der-Avakian, A., & Markou, A. (2012). The neurobiology of anhedonia and other reward-related deficits. *Trends in Neurosciences*, 35(1), 68–77.
<https://doi.org/10.1016/j.tins.2011.11.005>
- Dickerson, F., Jones-Brando, L., Ford, G., Genovese, G., Stallings, C., Origoni, A., O'Dushlaine, C., Katsafanas, E., Sweeney, K., Khushalani, S., & Yolken, R. (2019). Schizophrenia is Associated With an Aberrant Immune Response to Epstein–Barr Virus. *Schizophrenia Bulletin*, 45(5), 1112–1119.
<https://doi.org/10.1093/schbul/sby164>
- Dieleman, J. L., Baral, R., Birger, M., Bui, A. L., Bulchis, A., Chapin, A., Hamavid, H., Horst, C., Johnson, E. K., Joseph, J., Lavado, R., Lomsadze, L., Reynolds, A., Squires, E., Campbell, M., DeCenso, B., Dicker, D., Flaxman, A. D., Gabert, R., ...

- Murray, C. J. L. (2016). US Spending on Personal Health Care and Public Health, 1996-2013. *JAMA*, 316(24), 2627–2646. <https://doi.org/10.1001/jama.2016.16885>
- Dinga, R., Marquand, A. F., Veltman, D. J., Beekman, A. T. F., Schoevers, R. A., van Hemert, A. M., Penninx, B. W. J. H., & Schmaal, L. (2018). Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: A machine learning approach. *Translational Psychiatry*, 8(1), 1–11. <https://doi.org/10.1038/s41398-018-0289-1>
- Dinga, R., Schmaal, L., Penninx, B. W. J. H., van Tol, M. J., Veltman, D. J., van Velzen, L., Mennes, M., van der Wee, N. J. A., & Marquand, A. F. (2019). Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *NeuroImage: Clinical*, 22, 101796. <https://doi.org/10.1016/j.nicl.2019.101796>
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- Dwyer, D. B., Kalman, J. L., Budde, M., Kambeitz, J., Ruef, A., Antonucci, L. A., Kambeitz-Illankovic, L., Hasan, A., Kondofersky, I., Anderson-Schmidt, H., Gade, K., Reich-erkelenz, D., Adorjan, K., Senner, F., Schaupp, S., Andlauer, T. F. M., Comes, A. L., Schulte, E. C., Klöhn-Saghatolislam, F., ... Koutsouleris, N. (2020). An Investigation of Psychosis Subgroups With Prognostic Validation and Exploration of Genetic Underpinnings: The PsyCourse Study. *JAMA Psychiatry*, 77(5), 523–533. <https://doi.org/10.1001/jamapsychiatry.2019.4910>
- Ellison-Wright, I., Glahn, D. C., Laird, A. R., Thelen, S. M., & Bullmore, E. (2008). The Anatomy of First-Episode and Chronic Schizophrenia: An Anatomical Likelihood Estimation Meta-Analysis. *American Journal of Psychiatry*, 165(8), 1015–1023. <https://doi.org/10.1176/appi.ajp.2008.07101562>

- Endler, N. S. (2004). *CISS: Coping Inventory for Stressful Situations*. Harcourt.
- F. Schultze-Lutter, J. Addington, S. Ruhrmann, & J. Klosterkötter. (2007). *Schizophrenia Proneness Instrument, Adult Version (SPI-A)* (Giovanni Fioriti Editore).
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification. *J. Mach. Learn. Res.*, 9, 1871–1874.
- Feinberg, I. (1982). Schizophrenia: Caused by a fault in programmed synaptic elimination during adolescence? *Journal of Psychiatric Research*, 17(4), 319–334.
[https://doi.org/10.1016/0022-3956\(82\)90038-3](https://doi.org/10.1016/0022-3956(82)90038-3)
- Fillman, S. G., Cloonan, N., Catts, V. S., Miller, L. C., Wong, J., McCrossin, T., Cairns, M., & Weickert, C. S. (2013). Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Molecular Psychiatry*, 18(2), 206–214. <https://doi.org/10.1038/mp.2012.110>
- Fillman, S. G., Weickert, T. W., Lenroot, R. K., Catts, S. V., Bruggemann, J. M., Catts, V. S., & Weickert, C. S. (2016). Elevated peripheral cytokines characterize a subgroup of people with schizophrenia displaying poor verbal fluency and reduced Broca’s area volume. *Molecular Psychiatry*, 21(8), 1090–1098. <https://doi.org/10.1038/mp.2015.90>
- Filzmoser, P., Liebmann, B., & Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics*, 23(4), 160–171. <https://doi.org/10.1002/cem.1225>
- First, M. B., Rebello, T. J., Keeley, J. W., Bhargava, R., Dai, Y., Kulygina, M., Matsumoto, C., Robles, R., Stona, A., & Reed, G. M. (2018). Do mental health professionals use diagnostic classifications the way we think they do? A global survey. *World Psychiatry*, 17(2), 187–195. <https://doi.org/10.1002/wps.20525>
- First, Michael B, Spitzer, Robert L, Gibbon, Miriam, & Williams, Janet B.W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version*,

- Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. Biometrics Research, New York State Psychiatric Institute.
- Fischer, B. A., & Carpenter, W. T. (2009). Will The Kraepelinian Dichotomy Survive DSM-V? *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, 34(9), 2081–2087. <https://doi.org/10.1038/npp.2009.32>
- Fond, G., Lançon, C., Korchia, T., Auquier, P., & Boyer, L. (2020). The Role of Inflammation in the Treatment of Schizophrenia. *Frontiers in Psychiatry*, 11, 160. <https://doi.org/10.3389/fpsyt.2020.00160>
- Fornito, A., Yucel, M., Dean, B., Wood, S. J., & Pantelis, C. (2009). Anatomical Abnormalities of the Anterior Cingulate Cortex in Schizophrenia: Bridging the Gap Between Neuroimaging and Neuropathology. *Schizophrenia Bulletin*, 35(5), 973–993. <https://doi.org/10.1093/schbul/sbn025>
- Fornito, A., Yücel, M., Patti, J., Wood, S. J., & Pantelis, C. (2009). Mapping grey matter reductions in schizophrenia: An anatomical likelihood estimation analysis of voxel-based morphometry studies. *Schizophrenia Research*, 108(1–3), 104–113. <https://doi.org/10.1016/j.schres.2008.12.011>
- Foulds, G. A., & Bedford, A. (1975). Hierarchy of classes of personal illness. *Psychological Medicine*, 5(2), 181–192. <https://doi.org/10.1017/S0033291700056452>
- Friborg, O., Hjemdal, O., Rosenvinge, J. H., & Martinussen, M. (2003). A new rating scale for adult resilience: What are the central protective resources behind healthy adjustment? *International Journal of Methods in Psychiatric Research*, 12(2), 65–76.
- Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, 172, 96–102. <https://doi.org/10.1016/j.jad.2014.10.010>

- Fusar-Poli, P., Solmi, M., Brondino, N., Davies, C., Chae, C., Politi, P., Borgwardt, S., Lawrie, S. M., Parnas, J., & McGuire, P. (2019). Transdiagnostic psychiatry: A systematic review. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 18(2), 192–207. <https://doi.org/10.1002/wps.20631>
- Fusar-Poli, P., Stringer, D., M. S. Durieux, A., Rutigliano, G., Bonoldi, I., De Micheli, A., & Stahl, D. (2019). Clinical-learning versus machine-learning for transdiagnostic prediction of psychosis onset in individuals at-risk. *Translational Psychiatry*, 9, 259. <https://doi.org/10.1038/s41398-019-0600-9>
- Giedd, J. N., Keshavan, M., & Paus, T. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews. Neuroscience*, 9(12), 947–957. <https://doi.org/10.1038/nrn2513>
- Gogtay, N., Vyas, N. S., Testa, R., Wood, S. J., & Pantelis, C. (2011). Age of Onset of Schizophrenia: Perspectives From Structural Neuroimaging Studies. *Schizophrenia Bulletin*, 37(3), 504–513. <https://doi.org/10.1093/schbul/sbr030>
- Golland, P., & Fischl, B. (2003). Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies. In C. Taylor & J. A. Noble (Eds.), *Information Processing in Medical Imaging* (pp. 330–341). Springer Berlin Heidelberg.
- Gollub, R. L., Shoemaker, J. M., King, M. D., White, T., Ehrlich, S., Sponheim, S. R., Clark, V. P., Turner, J. A., Mueller, B. A., Magnotta, V., O’Leary, D., Ho, B. C., Brauns, S., Manoach, D. S., Seidman, L., Bustillo, J. R., Lauriello, J., Bockholt, J., Lim, K. O., ... Andreasen, N. C. (2013). The MCIC collection: A shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*, 11(3), 367–388. <https://doi.org/10.1007/s12021-013-9184-3>
- Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., Ortega, B. N., Zaiko, Y. V., Roach, E. L., Korgaonkar, M. S., Grieve, S. M., Galatzer-

- Levy, I., Fox, P. T., & Etkin, A. (2015). Identification of a Common Neurobiological Substrate for Mental Illness. *JAMA Psychiatry*, 72(4), 305–315.
<https://doi.org/10.1001/jamapsychiatry.2014.2206>
- Grace, A. A. (2012). Dopamine System Dysregulation by the Hippocampus: Implications for the Pathophysiology and Treatment of Schizophrenia. *Neuropharmacology*, 62(3), 1342–1348. <https://doi.org/10.1016/j.neuropharm.2011.05.011>
- Gray, J. P., Müller, V. I., Eickhoff, S. B., & Fox, P. T. (2020). Multimodal Abnormalities of Brain Structure and Function in Major Depressive Disorder: A Meta-Analysis of Neuroimaging Studies. *American Journal of Psychiatry*, 177(5), 422–434.
<https://doi.org/10.1176/appi.ajp.2019.19050560>
- Griffiths, S. L., Leighton, S. P., Mallikarjun, P. K., Blake, G., Everard, L., Jones, P. B., Fowler, D., Hodgekins, J., Amos, T., Freemantle, N., Sharma, V., Marshall, M., McCrone, P., Singh, S. P., Birchwood, M., & Upthegrove, R. (2021). Structure and stability of symptoms in first episode psychosis: A longitudinal network approach. *Translational Psychiatry*, 11(1), 1–8. <https://doi.org/10.1038/s41398-021-01687-y>
- Han, L. K. M., Dinga, R., Hahn, T., Ching, C. R. K., Eyler, L. T., Aftanas, L., Aghajani, M., Aleman, A., Baune, B. T., Berger, K., Brak, I., Filho, G. B., Carballedo, A., Connolly, C. G., Couvy-Duchesne, B., Cullen, K. R., Dannlowski, U., Davey, C. G., Dima, D., ... Schmaal, L. (2020). Brain aging in major depressive disorder: Results from the ENIGMA major depressive disorder working group. *Molecular Psychiatry*, 1–16.
<https://doi.org/10.1038/s41380-020-0754-0>
- Han, M.-H., & Nestler, E. J. (2017). Neural Substrates of Depression and Resilience. *Neurotherapeutics: The Journal of the American Society for Experimental Neurotherapeutics*, 14(3), 677–686. <https://doi.org/10.1007/s13311-017-0527-x>

- Harrow, M., Grossman, L. S., Herbener, E. S., & Davies, E. W. (2000). Ten-year outcome: Patients with schizoaffective disorders, schizophrenia, affective disorders and mood-incongruent psychotic symptoms. *British Journal of Psychiatry*, 177(5), 421–426. <https://doi.org/10.1192/bjp.177.5.421>
- Harvey, A. G., Murray, G., Chandler, R. A., & Soehner, A. (2011). Sleep disturbance as transdiagnostic: Consideration of neurobiological mechanisms. *Clinical Psychology Review*, 31(2), 225–235. <https://doi.org/10.1016/j.cpr.2010.04.003>
- Harvey, A. G., Watkins, E., & Mansell, W. (2004). *Cognitive Behavioural Processes Across Psychological Disorders: A Transdiagnostic Approach to Research and Treatment*. Oxford University Press.
- Hasler, G., Drevets, W. C., Manji, H. K., & Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 29(10), 1765–1781. <https://doi.org/10.1038/sj.npp.1300506>
- Haukvik, U. K., Tamnes, C. K., Söderman, E., & Agartz, I. (2018). Neuroimaging hippocampal subfields in schizophrenia and bipolar disorder: A systematic review and meta-analysis. *Journal of Psychiatric Research*, 104, 217–226. <https://doi.org/10.1016/j.jpsychires.2018.08.012>
- Health (UK), N. C. C. for M. (2014). *PSYCHOSIS AND SCHIZOPHRENIA IN ADULTS*. National Institute for Health and Care Excellence (UK). <https://www.ncbi.nlm.nih.gov/books/NBK333029/>
- Health, W. H. O. D. of M. (1996). *WHOQOL-BREF: Introduction, administration, scoring and generic version of the assessment : field trial version, December 1996* (WHOQOL-BREF). Article WHOQOL-BREF. <https://apps.who.int/iris/handle/10665/63529>

- Hengartner, M. P., & Lehmann, S. N. (2017). Why Psychiatric Research Must Abandon Traditional Diagnostic Classification and Adopt a Fully Dimensional Scope: Two Solutions to a Persistent Problem. *Frontiers in Psychiatry*, 8.
<https://doi.org/10.3389/fpsyt.2017.00101>
- Herniman, S. E., Phillips, L. J., Wood, S. J., Cotton, S. M., Liemburg, E. J., & Allott, K. A. (2021). Interrelationships between depressive symptoms and positive and negative symptoms of recent onset schizophrenia spectrum disorders: A network analytical approach. *Journal of Psychiatric Research*, 140, 373–380.
<https://doi.org/10.1016/j.jpsychires.2021.05.038>
- Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive Therapy and Research*, 36(5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>
- Hopper, K., Harrison, G., Janca, A., & Sartorius, N. (Eds.). (2007). *Recovery from schizophrenia: An international perspective: A report from the WHO Collaborative Project, the international study of schizophrenia* (pp. xix, 370). Oxford University Press.
- Huang, H., Liu, Y., Yuan, M., & Marron, J. S. (2015). Statistical Significance of Clustering using Soft Thresholding. *Journal of Computational and Graphical Statistics : A Joint Publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, 24(4), 975–993.
<https://doi.org/10.1080/10618600.2014.948179>
- Jacomb, I., Stanton, C., Vasudevan, R., Powell, H., O'Donnell, M., Lenroot, R., Bruggemann, J., Balzan, R., Galletly, C., Liu, D., Weickert, C. S., & Weickert, T. W. (2018). C-Reactive Protein: Higher During Acute Psychotic Episodes and Related to

- Cortical Thickness in Schizophrenia and Healthy Controls. *Frontiers in Immunology*, 9, 2230. <https://doi.org/10.3389/fimmu.2018.02230>
- Jiang, J., Zhao, Y.-J., Hu, X.-Y., Du, M.-Y., Chen, Z.-Q., Wu, M., Li, K.-M., Zhu, H.-Y., Kumar, P., & Gong, Q.-Y. (2017). Microstructural brain abnormalities in medication-free patients with major depressive disorder: A systematic review and meta-analysis of diffusion tensor imaging. *Journal of Psychiatry & Neuroscience: JPN*, 42(3), 150–163. <https://doi.org/10.1503/jpn.150341>
- Jiang, Y., Duan, M., Chen, X., Zhang, X., Gong, J., Dong, D., Li, H., Yi, Q., Wang, S., Wang, J., Luo, C., & Yao, D. (2019). Aberrant Prefrontal–Thalamic–Cerebellar Circuit in Schizophrenia and Depression: Evidence From a Possible Causal Connectivity. *International Journal of Neural Systems*, 29(05), 1850032. <https://doi.org/10.1142/S0129065718500326>
- Johnsen, E., Fathian, F., Kroken, R. A., Steen, V. M., Jørgensen, H. A., Gjestad, R., & Løberg, E.-M. (2016). The serum level of C-reactive protein (CRP) is associated with cognitive performance in acute phase psychosis. *BMC Psychiatry*, 16, 60. <https://doi.org/10.1186/s12888-016-0769-x>
- Johnson, J., Horwath, E., & Weissman, M. M. (1991). The Validity of Major Depression With Psychotic Features Based on a Community Study. *Archives of General Psychiatry*, 48(12), 1075–1081. <https://doi.org/10.1001/archpsyc.1991.01810360039006>
- Kaczurkin, A. N., Sotiras, A., Baller, E. B., Barzilay, R., Calkins, M. E., Chand, G. B., Cui, Z., Erus, G., Fan, Y., Gur, R. E., Gur, R. C., Moore, T. M., Roalf, D. R., Rosen, A. F. G., Ruparel, K., Shinohara, R. T., Varol, E., Wolf, D. H., Davatzikos, C., & Satterthwaite, T. D. (2020). Neurostructural Heterogeneity in Youth with

- Internalizing Symptoms. *Biological Psychiatry*, 87(5), 473–482.
<https://doi.org/10.1016/j.biopsych.2019.09.005>
- Kambeitz, J., Cabral, C., Sacchet, M. D., Gotlib, I. H., Zahn, R., Serpa, M. H., Walter, M., Falkai, P., & Koutsouleris, N. (2017). Detecting Neuroimaging Biomarkers for Depression: A Meta-analysis of Multivariate Pattern Recognition Studies. *Biological Psychiatry*, 82(5), 330–338. <https://doi.org/10.1016/j.biopsych.2016.10.028>
- Kambeitz, J., Kambeitz-Illankovic, L., Leucht, S., Wood, S., Davatzikos, C., Malchow, B., Falkai, P., & Koutsouleris, N. (2015). Detecting Neuroimaging Biomarkers for Schizophrenia: A Meta-Analysis of Multivariate Pattern Recognition Studies. *Neuropsychopharmacology*, 40(7), 1742–1751. <https://doi.org/10.1038/npp.2015.22>
- Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17(12), 1174–1179. <https://doi.org/10.1038/mp.2012.105>
- Karrer, T., Ds, B., B, D., O, G., A, A., R, J., Ar, L., Pt, F., Sb, E., O, G., G, V., B, T., & D, B. (2019). Brain-based ranking of cognitive domains to predict schizophrenia. *Human Brain Mapping*, 40(15), 4487–4507. <https://doi.org/10.1002/hbm.24716>
- Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2), 261–276.
- Keller, J., Shen, L., Gomez, R. G., Garrett, A., Solvason, H. B., Reiss, A., & Schatzberg, A. F. (2008). Hippocampal and Amygdalar Volumes in Psychotic and Nonpsychotic Unipolar Depression. *The American Journal of Psychiatry*, 165(7), 872–880.
<https://doi.org/10.1176/appi.ajp.2008.07081257>
- Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., Andreassen, O. A., Arango, C., Banaj, N., Bouix, S., Bousman, C. A., Brouwer, R. M., Bruggemann, J., Bustillo, J., Cahn, W., Calhoun, V., Cannon, D., Carr, V., Catts, S., ... Donohoe,

- G. (2018). Widespread white matter microstructural differences in schizophrenia across 4322 individuals: Results from the ENIGMA Schizophrenia DTI Working Group. *Molecular Psychiatry*, 23(5), 1261–1269.
<https://doi.org/10.1038/mp.2017.170>
- 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. <https://doi.org/10.1001/archgenpsychiatry.2011.60>
- Kendler, K. S. (1991). Mood-Incongruent Psychotic Affective Illness: A Historical and Empirical Review. *Archives of General Psychiatry*, 48(4), 362–369.
<https://doi.org/10.1001/archpsyc.1991.01810280078012>
- Kendler, K. S. (2013). What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Molecular Psychiatry*, 18(10), 1058–1066.
<https://doi.org/10.1038/mp.2013.50>
- Kendler, K. S. (2018). Classification of psychopathology: Conceptual and historical background. *World Psychiatry*, 17(3), 241–242. <https://doi.org/10.1002/wps.20549>
- Kendler, K. S. (2020). The Development of Kraepelin’s Concept of Dementia Praecox: A Close Reading of Relevant Texts. *JAMA Psychiatry*, 77(11), 1181–1187.
<https://doi.org/10.1001/jamapsychiatry.2020.1266>
- Kern, D. M., Cepeda, M. S., Defalco, F., & Etropolski, M. (2020). Treatment patterns and sequences of pharmacotherapy for patients diagnosed with depression in the United States: 2014 through 2019. *BMC Psychiatry*, 20. <https://doi.org/10.1186/s12888-019-2418-7>
- Keshavan, M. S., & Brady, R. (2011). Biomarkers in schizophrenia: We need to rebuild the Titanic. *World Psychiatry*, 10(1), 35–36.

- Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, 20(4), 359–364.
<https://doi.org/10.1097/YCO.0b013e32816ebc8c>
- Khandaker, G. M., Cousins, L., Deakin, J., Lennox, B. R., Yolken, R., & Jones, P. B. (2015). Inflammation and immunity in schizophrenia: Implications for pathophysiology and treatment. *The Lancet. Psychiatry*, 2(3), 258–270. [https://doi.org/10.1016/S2215-0366\(14\)00122-9](https://doi.org/10.1016/S2215-0366(14)00122-9)
- Khandaker, G. M., Pearson, R. M., Zammit, S., Lewis, G., & Jones, P. B. (2014). Association of Serum Interleukin 6 and C-Reactive Protein in Childhood With Depression and Psychosis in Young Adult Life. *JAMA Psychiatry*, 71(10), 1121–1128.
<https://doi.org/10.1001/jamapsychiatry.2014.1332>
- Kline, E. R., Seidman, L. J., Cornblatt, B. A., Woodberry, K. A., Bryant, C., Bearden, C. E., Cadenhead, K. S., Cannon, T. D., Mathalon, D. H., McGlashan, T. H., Perkins, D. O., Tsuang, M. T., Walker, E. F., Woods, S. W., & Addington, J. (2018). Depression and clinical high-risk states: Baseline presentation of depressed vs. non-depressed participants in the NAPLS-2 cohort. *Schizophrenia Research*, 192, 357–363.
<https://doi.org/10.1016/j.schres.2017.05.032>
- Klöppel, S., Abdulkadir, A., Jack, C. R., Koutsouleris, N., Mourão-Miranda, J., & Vemuri, P. (2012). Diagnostic neuroimaging across diseases. *NeuroImage*, 61(2), 457–463.
<https://doi.org/10.1016/j.neuroimage.2011.11.002>
- Klosterkötter, J., SCHULTZE-LUTTER, F., BECHDOLF, A., & RUHRMANN, S. (2011). Prediction and prevention of schizophrenia: What has been achieved and where to go next? *World Psychiatry*, 10(3), 165–174.

- Kochunov, P., Glahn, D. C., Rowland, L. M., Olvera, R. L., Winkler, A., Yang, Y. H., Sampath, H., Carpenter, W. T., Dugarrrila, R., Curran, J., Blangero, J., & Hong, L. E. (2013). Testing the hypothesis of accelerated cerebral white matter aging in schizophrenia and major depression. *Biological Psychiatry*, 73(5), 482–491. <https://doi.org/10.1016/j.biopsych.2012.10.002>
- Koreen, A. R., Siris, S. G., Chakos, M., Alvir, J., Mayerhoff, D., & Lieberman, J. (1993). Depression in first-episode schizophrenia. *The American Journal of Psychiatry*, 150(11), 1643–1648. <https://doi.org/10.1176/ajp.150.11.1643>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Koutsouleris, N., Gaser, C., Jäger, M., Bottlender, R., Frodl, T., Holzinger, S., Schmitt, G. J. E., Zetzsche, T., Burgermeister, B., Scheuerecker, J., Born, C., Reiser, M., Möller, H.-J., & Meisenzahl, E. M. (2008). Structural correlates of psychopathological symptom dimensions in schizophrenia: A voxel-based morphometric study. *NeuroImage*, 39(4), 1600–1612. <https://doi.org/10.1016/j.neuroimage.2007.10.029>
- Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., Schmidt, A., Gillam, J., Schultze-Lutter, F., Falkai, P., Reiser, M., Riecher-Rössler, A., Upthegrove, R., Hietala, J., Salokangas, R. K. R., ... Borgwardt, S. (2018a). Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With

Recent-Onset Depression. *JAMA Psychiatry*, 75(11), 1156–1172.

<https://doi.org/10.1001/jamapsychiatry.2018.2165>

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., Schmidt, A., Gillam, J., Schultze-Lutter, F., Falkai, P., Reiser, M., Riecher-Rössler, A., Upthegrove, R., Hietala, J., Salokangas, R. K. R., ... Borgwardt, S. (2018b). Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry*, 75(11), 1156–1172.

<https://doi.org/10.1001/jamapsychiatry.2018.2165>

Koutsouleris, N., Meisenzahl, E. M., Borgwardt, S., Riecher-Rössler, A., Frodl, T., Kambeitz, J., Köhler, Y., Falkai, P., Möller, H.-J., Reiser, M., & Davatzikos, C. (2015). Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain*, 138(7), 2059–2073.

<https://doi.org/10.1093/brain/awv111>

Kreutzer, J. S., DeLuca, J., & Caplan, B. (Eds.). (2011). WAIS-III. In *Encyclopedia of Clinical Neuropsychology* (pp. 2667–2667). Springer. https://doi.org/10.1007/978-0-387-79948-3_4123

Krueger, R. F., & Eaton, N. R. (2015). Transdiagnostic factors of mental disorders. *World Psychiatry*, 14(1), 27–29. <https://doi.org/10.1002/wps.20175>

Krynicky, C. R., Upthegrove, R., Deakin, J. F. W., & Barnes, T. R. E. (2018). The relationship between negative symptoms and depression in schizophrenia: A systematic review. *Acta Psychiatrica Scandinavica*, 137(5), 380–390.

<https://doi.org/10.1111/acps.12873>

- Kumari, V., Peters, E., Guinn, A., Fannon, D., Russell, T., Sumich, A., Kuipers, E., Williams, S. C. R., & ffytche, D. H. (2016). Mapping Depression in Schizophrenia: A Functional Magnetic Resonance Imaging Study. *Schizophrenia Bulletin*, 42(3), 802–813. <https://doi.org/10.1093/schbul/sbv186>
- Lalousis, P. A., Wood, S. J., Schmaal, L., Chisholm, K., Griffiths, S. L., Reniers, R. L. E. P., Bertolino, A., Borgwardt, S., Brambilla, P., Kambeitz, J., Lencer, R., Pantelis, C., Ruhrmann, S., Salokangas, R. K. R., Schultze-Lutter, F., Bonivento, C., Dwyer, D., Ferro, A., Haidl, T., ... PRONIA Consortium. (2021). Heterogeneity and Classification of Recent Onset Psychosis and Depression: A Multimodal Machine Learning Approach. *Schizophrenia Bulletin*, sbaa185. <https://doi.org/10.1093/schbul/sbaa185>
- Laskaris, L., Mancuso, S., Shannon Weickert, C., Zalesky, A., Chana, G., Wannan, C., Bousman, C., Baune, B. T., McGorry, P., Pantelis, C., & Cropley, V. L. (2021). Brain morphology is differentially impacted by peripheral cytokines in schizophrenia-spectrum disorder. *Brain, Behavior, and Immunity*, 95, 299–309. <https://doi.org/10.1016/j.bbi.2021.04.002>
- Lesh, T. A., Careaga, M., Rose, D. R., McAllister, A. K., Van de Water, J., Carter, C. S., & Ashwood, P. (2018). Cytokine alterations in first-episode schizophrenia and bipolar disorder: Relationships to brain structure and symptoms. *Journal of Neuroinflammation*, 15(1), 165. <https://doi.org/10.1186/s12974-018-1197-2>
- Li, H., Chang, H., Song, X., Liu, W., Li, L., Wang, L., Yang, Y., Zhang, L., Li, W., Zhang, Y., Zhou, D.-S., Li, X., Zhang, C., Fang, Y., Sun, Y., Dai, J.-P., Luo, X.-J., Yao, Y.-G., Xiao, X., ... Li, M. (2019). Integrative analyses of major histocompatibility complex loci in the genome-wide association studies of major depressive disorder.

Neuropsychopharmacology, 44(9), 1552–1561. <https://doi.org/10.1038/s41386-019-0346-3>

- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, 339, b2700. <https://doi.org/10.1136/bmj.b2700>
- Linden, D. E. J. (2012). The Challenges and Promise of Neuroimaging in Psychiatry. *Neuron*, 73(1), 8–22. <https://doi.org/10.1016/j.neuron.2011.12.014>
- Liu, Z., Palaniyappan, L., Wu, X., Zhang, K., Du, J., Zhao, Q., Xie, C., Tang, Y., Su, W., Wei, Y., Xue, K., Han, S., Tsai, S.-J., Lin, C.-P., Cheng, J., Li, C., Wang, J., Sahakian, B. J., Robbins, T. W., ... Feng, J. (2021). Resolving heterogeneity in schizophrenia through a novel systems approach to brain structure: Individualized structural covariance network analysis. *Molecular Psychiatry*, 1–13. <https://doi.org/10.1038/s41380-021-01229-4>
- Llorca, P.-M., Lançon, C., Lancrenon, S., Bayle, F.-J., Caci, H., Rouillon, F., & Gorwood, P. (2009). The “Functional Remission of General Schizophrenia” (FROGS) scale: Development and validation of a new questionnaire. *Schizophrenia Research*, 113(2–3), 218–225. <https://doi.org/10.1016/j.schres.2009.04.029>
- Loughland, C., Draganic, D., McCabe, K., Richards, J., Nasir, A., Allen, J., Catts, S., Jablensky, A., Henskens, F., Michie, P., Mowry, B., Pantelis, C., Schall, U., Scott, R., Tooney, P., & Carr, V. (2010). Australian Schizophrenia Research Bank: A Database of Comprehensive Clinical, Endophenotypic and Genetic Data for Aetiological Studies of Schizophrenia. *Australian & New Zealand Journal of Psychiatry*, 44(11), 1029–1035. <https://doi.org/10.3109/00048674.2010.501758>

- Luo, N., Tian, L., Calhoun, V. D., Chen, J., Lin, D., Vergara, V. M., Rao, S., Yang, J., Zhuo, C., Xu, Y., Turner, J. A., Zhang, F., & Sui, J. (2019). Brain function, structure and genomic data are linked but show different sensitivity to duration of illness and disease stage in schizophrenia. *NeuroImage: Clinical*, 23, 101887.
<https://doi.org/10.1016/j.nicl.2019.101887>
- Madre, M., Canales-Rodríguez, E. J., Ortiz-Gil, J., Murru, A., Torrent, C., Bramon, E., Perez, V., Orth, M., Brambilla, P., Vieta, E., & Amann, B. L. (2016). Neuropsychological and neuroimaging underpinnings of schizoaffective disorder: A systematic review. *Acta Psychiatrica Scandinavica*, 134(1), 16–30. <https://doi.org/10.1111/acps.12564>
- Maj, M. (2016). The need for a conceptual framework in psychiatry acknowledging complexity while avoiding defeatism. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, 15(1), 1–2. <https://doi.org/10.1002/wps.20291>
- Manjón, J. V., Tohka, J., García-Martí, G., Carbonell-Caballero, J., Lull, J. J., Martí-Bonmatí, L., & Robles, M. (2008). Robust MRI brain tissue parameter estimation by multistage outlier rejection. *Magnetic Resonance in Medicine*, 59(4), 866–873.
<https://doi.org/10.1002/mrm.21521>
- Marulanda, S., & Addington, J. (2016). Resilience in Individuals at Clinical High Risk for Psychosis. *Early Intervention in Psychiatry*, 10(3), 212–219.
<https://doi.org/10.1111/eip.12174>
- McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: A review and synthesis. *Neuropsychology*, 24(1), 9–34.
<https://doi.org/10.1037/a0017336>
- McGlashan, T. H., Miller, T. J., Woods, S. W., Hoffman, R. E., & Davidson, L. (2001). Instrument for the Assessment of Prodromal Symptoms and States. In T. Miller, S. A.

- Mednick, T. H. McGlashan, J. Libiger, & J. O. Johannessen (Eds.), *Early Intervention in Psychotic Disorders* (pp. 135–149). Springer Netherlands.
https://doi.org/10.1007/978-94-010-0892-1_7
- McGorry, P. D. (2013). The next stage for diagnosis: Validity through utility. *World Psychiatry*, 12(3), 213–215. <https://doi.org/10.1002/wps.20080>
- McGorry, P. D., KILLACKEY, E., & YUNG, A. (2008). Early intervention in psychosis: Concepts, evidence and future directions. *World Psychiatry*, 7(3), 148–156.
- McGorry, P. D., Purcell, R., Goldstone, S., & Amminger, G. P. (2011). Age of onset and timing of treatment for mental and substance use disorders: Implications for preventive intervention strategies and models of care. *Current Opinion in Psychiatry*, 24(4), 301–306. <https://doi.org/10.1097/YCO.0b013e3283477a09>
- McGorry, P., Keshavan, M., Goldstone, S., Amminger, P., Allott, K., Berk, M., Lavoie, S., Pantelis, C., Yung, A., Wood, S., & Hickie, I. (2014). Biomarkers and clinical staging in psychiatry. *World Psychiatry*, 13(3), 211–223. <https://doi.org/10.1002/wps.20144>
- McTeague, L. M., Rosenberg, B. M., Lopez, J. W., Carreon, D. M., Huemer, J., Jiang, Y., Chick, C. F., Eickhoff, S. B., & Etkin, A. (2020). Identification of Common Neural Circuit Disruptions in Emotional Processing Across Psychiatric Disorders. *The American Journal of Psychiatry*, 177(5), 411–421.
<https://doi.org/10.1176/appi.ajp.2019.18111271>
- Mednick, S., Huttunen, M. O., & Machón, R. A. (1994). Prenatal Influenza Infections and Adult Schizophrenia. *Schizophrenia Bulletin*, 20(2), 263–267.
<https://doi.org/10.1093/schbul/20.2.263>
- Meisenzahl, E. M., Rujescu, D., Kirner, A., Giegling, I., Kathmann, N., Leinsinger, G., Maag, K., Hegerl, U., Hahn, K., & Möller, H.-J. (2001). Association of an Interleukin-1 β Genetic Polymorphism With Altered Brain Structure in Patients With

- Schizophrenia. *American Journal of Psychiatry*, 158(8), 1316–1319.
<https://doi.org/10.1176/appi.ajp.158.8.1316>
- Meisenzahl, E. M., Seifert, D., Bottlender, R., Teipel, S., Zetzsche, T., Jäger, M., Koutsouleris, N., Schmitt, G., Scheuerecker, J., Burgermeister, B., Hampel, H., Rupprecht, T., Born, C., Reiser, M., Möller, H.-J., & Frodl, T. (2010). Differences in hippocampal volume between major depression and schizophrenia: A comparative neuroimaging study. *European Archives of Psychiatry and Clinical Neuroscience*, 260(2), 127–137. <https://doi.org/10.1007/s00406-009-0023-3>
- Ménard, C., Hodes, G. E., & Russo, S. J. (2016). Pathogenesis of depression: Insights from human and rodent studies. *Neuroscience*, 321, 138–162.
<https://doi.org/10.1016/j.neuroscience.2015.05.053>
- Metcalf, S. A., Jones, P. B., Nordstrom, T., Timonen, M., Mäki, P., Miettunen, J., Jääskeläinen, E., Järvelin, M.-R., Stochl, J., Murray, G. K., Veijola, J., & Khandaker, G. M. (2017). Serum C-reactive protein in adolescence and risk of schizophrenia in adulthood: A prospective birth cohort study. *Brain, Behavior, and Immunity*, 59, 253–259. <https://doi.org/10.1016/j.bbi.2016.09.008>
- Milaneschi, Y., Kappelmann, N., Ye, Z., Lamers, F., Moser, S., Jones, P. B., Burgess, S., Penninx, B. W. J. H., & Khandaker, G. M. (2021). Association of inflammation with depression and anxiety: Evidence for symptom-specificity and potential causality from UK Biobank and NESDA cohorts. *Molecular Psychiatry*.
<https://doi.org/10.1038/s41380-021-01188-w>
- Milaneschi, Y., Lamers, F., Berk, M., & Penninx, B. W. J. H. (2020). Depression Heterogeneity and Its Biological Underpinnings: Toward Immunometabolic Depression. *Biological Psychiatry*, 88(5), 369–380.
<https://doi.org/10.1016/j.biopsych.2020.01.014>

- Miller, A. H., Maletic, V., & Raison, C. L. (2009). Inflammation and Its Discontents: The Role of Cytokines in the Pathophysiology of Major Depression. *Biological Psychiatry*, 65(9), 732–741. <https://doi.org/10.1016/j.biopsych.2008.11.029>
- Miller, B. J., Buckley, P., Seabolt, W., Mellor, A., & Kirkpatrick, B. (2011). Meta-analysis of cytokine alterations in schizophrenia: Clinical status and antipsychotic effects. *Biological Psychiatry*, 70(7), 663–671. <https://doi.org/10.1016/j.biopsych.2011.04.013>
- Miller, B. J., Culpepper, N., & Rapaport, M. H. (2014). C-reactive protein levels in schizophrenia: A review and meta-analysis. *Clinical Schizophrenia & Related Psychoses*, 7(4), 223–230. <https://doi.org/10.3371/CSRP.MICU.020813>
- Miller, B. J., & Goldsmith, D. R. (2019). Inflammatory biomarkers in schizophrenia: Implications for heterogeneity and neurobiology. *Biomarkers in Neuropsychiatry*, 1, 100006. <https://doi.org/10.1016/j.bionps.2019.100006>
- Misiak, B., Stańczykiewicz, B., Kotowicz, K., Rybakowski, J. K., Samochowiec, J., & Frydecka, D. (2018). Cytokines and C-reactive protein alterations with respect to cognitive impairment in schizophrenia and bipolar disorder: A systematic review. *Schizophrenia Research*, 192, 16–29. <https://doi.org/10.1016/j.schres.2017.04.015>
- Mitelman, S. A., & Buchsbaum, M. S. (2007). Very poor outcome schizophrenia: Clinical and neuroimaging aspects. *International Review of Psychiatry (Abingdon, England)*, 19(4), 345–357. <https://doi.org/10.1080/09540260701486563>
- Moher, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, 151(4), 264. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

- Mokhtari, R., & Lachman, H. M. (2016). The Major Histocompatibility Complex (MHC) in Schizophrenia: A Review. *Journal of Clinical & Cellular Immunology*, 7(6), 479. <https://doi.org/10.4172/2155-9899.1000479>
- Moritz, S., Birkner, C., Kloss, M., Jahn, H., Hand, I., Haasen, C., & Krausz, M. (2002). Executive functioning in obsessive–compulsive disorder, unipolar depression, and schizophrenia. *Archives of Clinical Neuropsychology*, 7.
- Mufford, M. S., Stein, D. J., Dalvie, S., Groenewold, N. A., Thompson, P. M., & Jahanshad, N. (2017). Neuroimaging genomics in psychiatry—A translational approach. *Genome Medicine*, 9(1), 102. <https://doi.org/10.1186/s13073-017-0496-z>
- Müller, N., Weidinger, E., Leitner, B., & Schwarz, M. J. (2015). The role of inflammation in schizophrenia. *Frontiers in Neuroscience*, 9. <https://doi.org/10.3389/fnins.2015.00372>
- Murphy, B. P., & Brewer, W. J. (2011). Early intervention in psychosis: Strengths and limitations of services. *Advances in Psychiatric Treatment*, 17(6), 401–407. <https://doi.org/10.1192/apt.bp.110.008573>
- Murray, R. M., & Lewis, S. W. (1987). Is schizophrenia a neurodevelopmental disorder? *British Medical Journal (Clinical Research Ed.)*, 295(6600), 681–682.
- Mushquash, C., & O'Connor, B. P. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods*, 38(3), 542–547. <https://doi.org/10.3758/BF03192810>
- Newman, D. L., Moffitt, T. E., Caspi, A., & Silva, P. A. (1998). Comorbid mental disorders: Implications for treatment and sample selection. *Journal of Abnormal Psychology*, 107(2), 305–311. <https://doi.org/10.1037//0021-843x.107.2.305>
- Nichols, T. E., & Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1), 1–25. <https://doi.org/10.1002/hbm.1058>

- North, H. F., Bruggemann, J., Cropley, V., Swaminathan, V., Sundram, S., Lenroot, R., Pereira, A. M., Zalesky, A., Bousman, C., Pantelis, C., Weickert, T. W., & Shannon Weickert, C. (2021). Increased peripheral inflammation in schizophrenia is associated with worse cognitive performance and related cortical thickness reductions. *European Archives of Psychiatry and Clinical Neuroscience*, 271(4), 595–607.
<https://doi.org/10.1007/s00406-021-01237-z>
- Noto, C., Maes, M., Ota, V. K., Teixeira, A. L., Bressan, R. A., Gadelha, A., & Brietzke, E. (2015). High predictive value of immune-inflammatory biomarkers for schizophrenia diagnosis and association with treatment resistance. *The World Journal of Biological Psychiatry*, 16(6), 422–429. <https://doi.org/10.3109/15622975.2015.1062552>
- Nowicki, S., & Duke, M. P. (2001). Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA). In *Interpersonal sensitivity: Theory and measurement* (pp. 183–198). Lawrence Erlbaum Associates Publishers.
- O'Connor, S., & Agius, M. (2015). A systematic review of structural and functional MRI differences between psychotic and nonpsychotic depression. *Psychiatria Danubina*, 27 Suppl 1, S235-239.
- Ohayon, M. M., & Schatzberg, A. F. (2002). Prevalence of Depressive Episodes With Psychotic Features in the General Population. *American Journal of Psychiatry*, 159(11), 1855–1861. <https://doi.org/10.1176/appi.ajp.159.11.1855>
- Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.-U., & Jönsson, B. (2012). The economic cost of brain disorders in Europe. *European Journal of Neurology*, 19(1), 155–162. <https://doi.org/10.1111/j.1468-1331.2011.03590.x>
- Osimo, E. F., Baxter, L. J., Lewis, G., Jones, P. B., & Khandaker, G. M. (2019). Prevalence of low-grade inflammation in depression: A systematic review and meta-analysis of

- CRP levels. *Psychological Medicine*, 49(12), 1958–1970.
<https://doi.org/10.1017/S0033291719001454>
- Osimo, E. F., Pillinger, T., Rodriguez, I. M., Khandaker, G. M., Pariante, C. M., & Howes, O. D. (2020). Inflammatory markers in depression: A meta-analysis of mean differences and variability in 5,166 patients and 5,083 controls. *Brain, Behavior, and Immunity*, 87, 901–909. <https://doi.org/10.1016/j.bbi.2020.02.010>
- Ota, M., Ishikawa, M., Sato, N., Hori, H., Sasayama, D., Hattori, K., Teraishi, T., Noda, T., Obu, S., Nakata, Y., Higuchi, T., & Kunugi, H. (2013). Discrimination between schizophrenia and major depressive disorder by magnetic resonance imaging of the female brain. *Journal of Psychiatric Research*, 47(10), 1383–1388.
<https://doi.org/10.1016/j.jpsychires.2013.06.010>
- Owen, M. J., Sawa, A., & Mortensen, P. B. (2016). Schizophrenia. *The Lancet*, 388(10039), 86–97. [https://doi.org/10.1016/S0140-6736\(15\)01121-6](https://doi.org/10.1016/S0140-6736(15)01121-6)
- Pantelis, C., Yücel, M., Bora, E., Fornito, A., Testa, R., Brewer, W. J., Velakoulis, D., & Wood, S. J. (2009). Neurobiological Markers of Illness Onset in Psychosis and Schizophrenia: The Search for a Moving Target. *Neuropsychology Review*, 19(3), 385–398. <https://doi.org/10.1007/s11065-009-9114-1>
- Peralta, V., & Cuesta, M. J. (2003). The nosology of psychotic disorders: A comparison among competing classification systems. *Schizophrenia Bulletin*, 29(3), 413–425.
<https://doi.org/10.1093/oxfordjournals.schbul.a007016>
- Perkins, D. O., Jeffries, C. D., Addington, J., Bearden, C. E., Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A., Mathalon, D. H., McGlashan, T. H., Seidman, L. J., Tsuang, M. T., Walker, E. F., Woods, S. W., & Heinssen, R. (2015). Towards a Psychosis Risk Blood Diagnostic for Persons Experiencing High-Risk Symptoms: Preliminary

- Results From the NAPLS Project. *Schizophrenia Bulletin*, 41(2), 419–428.
<https://doi.org/10.1093/schbul/sbu099>
- Perry, B. I., Upthegrove, R., Kappelmann, N., Jones, P. B., Burgess, S., & Khandaker, G. M. (2021). Associations of immunological proteins/traits with schizophrenia, major depression and bipolar disorder: A bi-directional two-sample mendelian randomization study. *Brain, Behavior, and Immunity*.
<https://doi.org/10.1016/j.bbi.2021.07.009>
- Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, 20(3), 249–262.
[https://doi.org/10.1016/0028-3932\(82\)90100-2](https://doi.org/10.1016/0028-3932(82)90100-2)
- Pillinger, T., Osimo, E. F., Brugger, S., Mondelli, V., McCutcheon, R. A., & Howes, O. D. (2019). A Meta-analysis of Immune Parameters, Variability, and Assessment of Modal Distribution in Psychosis and Test of the Immune Subgroup Hypothesis. *Schizophrenia Bulletin*, 45(5), 1120–1133. <https://doi.org/10.1093/schbul/sby160>
- Pope, M. A., Joober, R., & Malla, A. K. (2013). Diagnostic stability of first-episode psychotic disorders and persistence of comorbid psychiatric disorders over 1 year. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, 58(10), 588–594.
<https://doi.org/10.1177/070674371305801008>
- Potuzak, M., Ravichandran, C., Lewandowski, K. E., Ongür, D., & Cohen, B. M. (2012). Categorical vs dimensional classifications of psychotic disorders. *Comprehensive Psychiatry*, 53(8), 1118–1129. <https://doi.org/10.1016/j.comppsy.2012.04.010>
- Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *The Lancet*, 370(9590), 859–877.
[https://doi.org/10.1016/S0140-6736\(07\)61238-0](https://doi.org/10.1016/S0140-6736(07)61238-0)

Psychiatrie P und P und N (DGPPN) Deutsche Gesellschaft für. (2006). *DGPPN S3*

Treatment Guideline Schizophrenia and Psychotic Disorders. AWMF.

Purves-Tyson, T. D., Robinson, K., Brown, A. M., Boerrigter, D., Cai, H. Q., Weissleder, C.,

Owens, S. J., Rothmond, D. A., & Shannon Weickert, C. (2020). Increased

Macrophages and C1qA, C3, C4 Transcripts in the Midbrain of People With

Schizophrenia. *Frontiers in Immunology*, 11.

<https://www.frontiersin.org/article/10.3389/fimmu.2020.02002>

Quattrone, D., Di Forti, M., Gayer-Anderson, C., Ferraro, L., Jongsma, H. E., Tripoli, G., La

Cascia, C., La Barbera, D., Tarricone, I., Berardi, D., Szöke, A., Arango, C., Lasalvia,

A., Tortelli, A., Llorca, P.-M., de Haan, L., Velthorst, E., Bobes, J., Bernardo, M., ...

Reininghaus, U. (2019). Transdiagnostic dimensions of psychopathology at first

episode psychosis: Findings from the multinational EU-GEI study. *Psychological*

Medicine, 49(08), 1378–1391. <https://doi.org/10.1017/S0033291718002131>

R. Polikar. (2006). *Ensemble based systems in decision making—IEEE Journals & Magazine*.

<https://ieeexplore.ieee.org/document/1688199>

Radua, J., Ramella-Cravaro, V., Ioannidis, J. P. A., Reichenberg, A., Phiphopthatsanee, N.,

Amir, T., Yenn Thoo, H., Oliver, D., Davies, C., Morgan, C., McGuire, P., Murray, R.

M., & Fusar-Poli, P. (2018). What causes psychosis? An umbrella review of risk and

protective factors. *World Psychiatry*, 17(1), 49–66. <https://doi.org/10.1002/wps.20490>

Raison, C. L., Capuron, L., & Miller, A. H. (2006). Cytokines sing the blues: Inflammation

and the pathogenesis of depression. *Trends in Immunology*, 27(1), 24–31.

<https://doi.org/10.1016/j.it.2005.11.006>

Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation

of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*,

16(2), 176–186. <https://doi.org/10.1109/42.563663>

- Rajkowska, G., Miguel-Hidalgo, J. J., Wei, J., Dilley, G., Pittman, S. D., Meltzer, H. Y., Overholser, J. C., Roth, B. L., & Stockmeier, C. A. (1999). Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biological Psychiatry*, 45(9), 1085–1098. [https://doi.org/10.1016/s0006-3223\(99\)00041-4](https://doi.org/10.1016/s0006-3223(99)00041-4)
- Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity. *Journal of Clinical and Experimental Neuropsychology*, 20(3), 310–319. <https://doi.org/10.1076/jcen.20.3.310.823>
- Rehm, J., & Shield, K. D. (2019). Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Current Psychiatry Reports*, 21(2), 10. <https://doi.org/10.1007/s11920-019-0997-0>
- Reininghaus, U., Böhnke, J. R., Chavez-Baldini, U., Gibbons, R., Ivleva, E., Clementz, B. A., Pearlson, G. D., Keshavan, M. S., Sweeney, J. A., & Tamminga, C. A. (2019). Transdiagnostic dimensions of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 18(1), 67–76. <https://doi.org/10.1002/wps.20607>
- Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage: *Perceptual and Motor Skills*, 8(3). <https://doi.org/10.2466/pms.1958.8.3.271>
- Reniers, R. L. E. P., Lin, A., Yung, A. R., Koutsouleris, N., Nelson, B., Cropley, V. L., Velakoulis, D., McGorry, P. D., Pantelis, C., & Wood, S. J. (2017). Neuroanatomical Predictors of Functional Outcome in Individuals at Ultra-High Risk for Psychosis. *Schizophrenia Bulletin*, 43(2), 449–458. <https://doi.org/10.1093/schbul/sbw086>
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.). [The psychological examination in cases of traumatic encephalopathy. Problems.]. *Archives de Psychologie*, 28, 215–285.

- Rickwood, D., Van Dyke, N., & Telford, N. (2015). Innovation in youth mental health services in Australia: Common characteristics across the first headspace centres. *Early Intervention in Psychiatry*, 9(1), 29–37. <https://doi.org/10.1111/eip.12071>
- Robert L. Brennan. (2011). *Generalizability Theory and Classical Test Theory: Applied Measurement in Education: Vol 24, No 1*.
<https://www.tandfonline.com/doi/abs/10.1080/08957347.2011.532417>
- Roiser, J. P., Stephan, K. E., den Ouden, H. E. M., Friston, K. J., & Joyce, E. M. (2010). Adaptive and aberrant reward prediction signals in the human brain. *Neuroimage*, 50(2), 657–664. <https://doi.org/10.1016/j.neuroimage.2009.11.075>
- Romm, K. L., Rossberg, J. I., Hansen, C. F., Haug, E., Andreassen, O. A., & Melle, I. (2011). Self-esteem is associated with premorbid adjustment and positive psychotic symptoms in early psychosis. *BMC Psychiatry*, 11(1), 136.
<https://doi.org/10.1186/1471-244X-11-136>
- Rothschild, A. J., Winer, J., Flint, A. J., Mulsant, B. H., Whyte, E. M., Heo, M., Fratoni, S., Gabriele, M., Kasapinovic, S., Meyers, B. S., & Study of Pharmacotherapy of Psychotic Depression (STOP-PD) Collaborative Study Group. (2008). Missed diagnosis of psychotic depression at 4 academic medical centers. *The Journal of Clinical Psychiatry*, 69(8), 1293–1296.
- Russell, S. J., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice Hall.
- Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., & Engemann, D. A. (2020). Predictive regression modeling with MEG/EEG: From source power to signals and cognitive states. *NeuroImage*, 222, 116893. <https://doi.org/10.1016/j.neuroimage.2020.116893>
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507–2517.
<https://doi.org/10.1093/bioinformatics/btm344>

- Salvatore, P., Baldessarini, R. J., Khalsa, H.-M. K., Amore, M., Vittorio, C. D., Ferraro, G., Maggini, C., & Tohen, M. (2013). Predicting Diagnostic Change Among Patients Diagnosed With First-Episode DSM-IV-TR Major Depressive Disorder With Psychotic Features. *The Journal of Clinical Psychiatry*, 74(7), 723–731. <https://doi.org/10.4088/JCP.12m08328>
- Sánchez-Torres, A. M., Elosúa, M. R., Lorente-Omeñaca, R., Moreno-Izco, L., & Cuesta, M. J. (2015). A comparative study of the working memory multicomponent model in psychosis and healthy controls. *Comprehensive Psychiatry*, 61, 97–105. <https://doi.org/10.1016/j.comppsy.2015.05.008>
- Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., Cheung, J. W., van Erp, T. G. M., Bos, D., Ikram, M. A., Vernooij, M. W., Niessen, W. J., Tiemeier, H., Hofman, A., Wittfeld, K., Grabe, H. J., Janowitz, D., Bülow, R., Selonke, M., ... 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. <https://doi.org/10.1038/mp.2016.60>
- Schmaal, L., Veltman, D. J., van Erp, T. G. M., Sämann, P. G., Frodl, T., Jahanshad, N., Loehrer, E., Tiemeier, H., Hofman, A., Niessen, W. J., Vernooij, M. W., Ikram, M. A., Wittfeld, K., Grabe, H. J., Block, A., Hegenscheid, K., Völzke, H., Hoehn, D., Czisch, M., ... 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. <https://doi.org/10.1038/mp.2015.69>
- Schnack, H. G., van Haren, N. E. M., Nieuwenhuis, M., Hulshoff Pol, H. E., Cahn, W., & Kahn, R. S. (2016). Accelerated Brain Aging in Schizophrenia: A Longitudinal

- Pattern Recognition Study. *The American Journal of Psychiatry*, 173(6), 607–616.
<https://doi.org/10.1176/appi.ajp.2015.15070922>
- Schultze-Lutter, F., Ruhrmann, S., Picker, H., von Reventlow, H. G., Brockhaus-Dumke, A., & Klosterkötter, J. (2007). Basic symptoms in early psychotic and depressive disorders. *The British Journal of Psychiatry. Supplement*, 51, s31-37.
<https://doi.org/10.1192/bjp.191.51.s31>
- Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose, S. A., Handsaker, R. E., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Daly, M. J., Carroll, M. C., Stevens, B., & McCarroll, S. A. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, 530(7589), 177–183.
<https://doi.org/10.1038/nature16549>
- Shankman, S. A., Nelson, B. D., Harrow, M., & Faull, R. (2010). Does physical anhedonia play a role in depression? A 20-year longitudinal study. *Journal of Affective Disorders*, 120(1–3), 170–176. <https://doi.org/10.1016/j.jad.2009.05.002>
- Shenton, M. E., Dickey, C. C., Frumin, M., & McCarley, R. W. (2001). A review of MRI findings in schizophrenia. *Schizophrenia Research*, 49(1–2), 1–52.
- Silverstone, P. H., & Salsali, M. (2003). Low self-esteem and psychiatric patients: Part I – The relationship between low self-esteem and psychiatric diagnosis. *Annals of General Hospital Psychiatry*, 2(1), 2. <https://doi.org/10.1186/1475-2832-2-2>
- Slavich, G. M., & Irwin, M. R. (2014). From Stress to Inflammation and Major Depressive Disorder: A Social Signal Transduction Theory of Depression. *Psychological Bulletin*, 140(3), 774–815. <https://doi.org/10.1037/a0035302>
- Smeland, O. B., Bahrami, S., Frei, O., Shadrin, A., O’Connell, K., Savage, J., Watanabe, K., Krull, F., Bettella, F., Steen, N. E., Ueland, T., Posthuma, D., Djurovic, S., Dale, A.

- M., & Andreassen, O. A. (2019). Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Molecular Psychiatry*. <https://doi.org/10.1038/s41380-018-0332-x>
- Sobocki, P., Jönsson, B., Angst, J., & Rehnberg, C. (2006). Cost of Depression in Europe. *Journal of Mental Health Policy and Economics*, 9(2), 87–98.
- Sönmez, N., Romm, K. L., Andreassen, O. A., Melle, I., & Røssberg, J. I. (2013). Depressive symptoms in first episode psychosis: A one-year follow-up study. *BMC Psychiatry*, 13(1), 106. <https://doi.org/10.1186/1471-244X-13-106>
- Steen, R. G., Mull, C., McClure, R., Hamer, R. M., & Lieberman, J. A. (2006). Brain volume in first-episode schizophrenia: Systematic review and meta-analysis of magnetic resonance imaging studies. *The British Journal of Psychiatry: The Journal of Mental Science*, 188, 510–518. <https://doi.org/10.1192/bjp.188.6.510>
- Stein, D. J., Phillips, K. A., Bolton, D., Fulford, K. W. M., Sadler, J. Z., & Kendler, K. S. (2010). What is a Mental/Psychiatric Disorder? From DSM-IV to DSM-V. *Psychological Medicine*, 40(11), 1759–1765. <https://doi.org/10.1017/S0033291709992261>
- Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., Heinz, A., Huys, Q. J. M., Owen, M. J., Binder, E. B., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Montague, P. R., Schnyder, U., Wang, X.-J., & Breakspear, M. (2016). Charting the landscape of priority problems in psychiatry, part 1: Classification and diagnosis. *The Lancet. Psychiatry*, 3(1), 77–83. [https://doi.org/10.1016/S2215-0366\(15\)00361-2](https://doi.org/10.1016/S2215-0366(15)00361-2)
- Stephan, K. E., Binder, E. B., Breakspear, M., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Schnyder, U., Wang, X.-J., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Heinz, A., Huys, Q. J. M., Montague, P. R., Owen, M. J., & Friston, K.

- J. (2016). Charting the landscape of priority problems in psychiatry, part 2: Pathogenesis and aetiology. *The Lancet. Psychiatry*, 3(1), 84–90.
[https://doi.org/10.1016/S2215-0366\(15\)00360-0](https://doi.org/10.1016/S2215-0366(15)00360-0)
- Tamminga, C. A., Ivleva, E. I., Keshavan, M. S., Pearlson, G. D., Clementz, B. A., Witte, B., Morris, D. W., Bishop, J., Thaker, G. K., & Sweeney, J. A. (2013). Clinical phenotypes of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *The American Journal of Psychiatry*, 170(11), 1263–1274.
<https://doi.org/10.1176/appi.ajp.2013.12101339>
- Tamouza, R., Meyer, U., Foiselle, M., Richard, J.-R., Wu, C.-L., Boukouaci, W., Le Corvoisier, P., Barrau, C., Lucas, A., Perron, H., & Leboyer, M. (2021). Identification of inflammatory subgroups of schizophrenia and bipolar disorder patients with HERV-W ENV antigenemia by unsupervised cluster analysis. *Translational Psychiatry*, 11(1), 1–8. <https://doi.org/10.1038/s41398-021-01499-0>
- The Cancer Genome Atlas Research Network. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *New England Journal of Medicine*, 372(26), 2481–2498. <https://doi.org/10.1056/NEJMoa1402121>
- The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. (1992). <https://apps.who.int/iris/handle/10665/37958>
- Toenders, Y. J., Laskaris, L., Davey, C. G., Berk, M., Milaneschi, Y., Lamers, F., Penninx, B. W. J. H., & Schmaal, L. (2021). Inflammation and depression in young people: A systematic review and proposed inflammatory pathways. *Molecular Psychiatry*.
<https://doi.org/10.1038/s41380-021-01306-8>
- Tonna, M., De Panfilis, C., & Marchesi, C. (2012). Mood-congruent and mood-incongruent psychotic symptoms in major depression: The role of severity and personality.

Journal of Affective Disorders, 141(2–3), 464–468.

<https://doi.org/10.1016/j.jad.2012.03.017>

Upthegrove, R., & Barnes, N. M. (2014). The immune system and schizophrenia: An update for clinicians. *Advances in Psychiatric Treatment*, 20(2), 83–91.

<https://doi.org/10.1192/apt.bp.113.011452>

Upthegrove, R., Lalousis, P., Mallikarjun, P., Chisholm, K., Griffiths, S. L., Iqbal, M., Pelton, M., Reniers, R., Stainton, A., Rosen, M., Ruef, A., Dwyer, D. B., Surman, M., Haidl, T., Penzel, N., Kambeitz-Illankovic, L., Bertolino, A., Brambilla, P., Borgwardt, S., ... PRONIA Consortium, P. (n.d.). The Psychopathology and Neuroanatomical Markers of Depression in Early Psychosis. *Schizophrenia Bulletin*.

<https://doi.org/10.1093/schbul/sbaa094>

Upthegrove, R., Lalousis, P., Mallikarjun, P., Chisholm, K., Griffiths, S. L., Iqbal, M., Pelton, M., Reniers, R., Stainton, A., Rosen, M., Ruef, A., Dwyer, D. B., Surman, M., Haidl, T., Penzel, N., Kambeitz-Illankovic, L., Bertolino, A., Brambilla, P., Borgwardt, S., ... PRONIA Consortium. (2020). The Psychopathology and Neuroanatomical Markers of Depression in Early Psychosis. *Schizophrenia Bulletin*, sbaa094.

<https://doi.org/10.1093/schbul/sbaa094>

Upthegrove, R., Manzanares-Teson, N., & Barnes, N. M. (2014). Cytokine function in medication-naïve first episode psychosis: A systematic review and meta-analysis. *Schizophrenia Research*, 155(1), 101–108.

<https://doi.org/10.1016/j.schres.2014.03.005>

Upthegrove, R., Marwaha, S., & Birchwood, M. (2017). Depression and Schizophrenia: Cause, Consequence, or Trans-diagnostic Issue? *Schizophrenia Bulletin*, 43(2), 240–244. <https://doi.org/10.1093/schbul/sbw097>

- van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, 27(6), 759–773. <https://doi.org/10.1177/0959354317737185>
- van Erp, T. G.M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., Pearlson, G. D., Yao, N., Fukunaga, M., Hashimoto, R., Okada, N., Yamamori, H., Bustillo, J. R., Clark, V. P., Agartz, I., Mueller, B. A., Cahn, W., de Zwarte, S. MC., Hulshoff Pol, H. E., ... Turner, J. A. (2018). Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 controls via the ENIGMA consortium. *Biological Psychiatry*, 84(9), 644–654. <https://doi.org/10.1016/j.biopsych.2018.04.023>
- Van Os, J., Gilvarry, C., Bale, R., Van Horn, E., Tattan, T., White, I., & Murray On Behalf Of The Uk700 Group, R. (1999). A comparison of the utility of dimensional and categorical representations of psychosis. *Psychological Medicine*, 29(3), 595–606. <https://doi.org/10.1017/S0033291798008162>
- van Os, J., Gilvarry, C., Bale, R., van Horn, E., Tattan, T., White, I., & Murray, R. (2000). Diagnostic value of the DSM and ICD categories of psychosis: An evidence-based approach. *Social Psychiatry and Psychiatric Epidemiology*, 35(7), 305–311. <https://doi.org/10.1007/s001270050243>
- Van Rheenen, T. E., Cropley, V., Zalesky, A., Bousman, C., Wells, R., Bruggemann, J., Sundram, S., Weinberg, D., Lenroot, R. K., Pereira, A., Shannon Weickert, C., Weickert, T. W., & Pantelis, C. (2018). Widespread Volumetric Reductions in Schizophrenia and Schizoaffective Patients Displaying Compromised Cognitive Abilities. *Schizophrenia Bulletin*, 44(3), 560–574. <https://doi.org/10.1093/schbul/sbx109>
- van Velzen, L. S., Kelly, S., Isaev, D., Aleman, A., Aftanas, L. I., Bauer, J., Baune, B. T., Brak, I. V., Carballido, A., Connolly, C. G., Couvy-Duchesne, B., Cullen, K. R.,

- Danilenko, K. V., Dannlowski, U., Enneking, V., Filimonova, E., Förster, K., Frodl, T., Gotlib, I. H., ... Schmaal, L. (2020). White matter disturbances in major depressive disorder: A coordinated analysis across 20 international cohorts in the ENIGMA MDD working group. *Molecular Psychiatry*, 25(7), 1511–1525.
<https://doi.org/10.1038/s41380-019-0477-2>
- Varol, E., Sotiras, A., & Davatzikos, C. (2017). HYDRA: Revealing Heterogeneity of imaging and genetic patterns through a multiple max-margin Discriminative Analysis framework. *NeuroImage*, 145(Pt B), 346–364.
<https://doi.org/10.1016/j.neuroimage.2016.02.041>
- Veale, J. F. (2014). Edinburgh Handedness Inventory—Short Form: A revised version based on confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, 19(2), 164–177. <https://doi.org/10.1080/1357650X.2013.783045>
- Velakoulis, D., Pantelis, C., McGorry, P. D., Dudgeon, P., Brewer, W., Cook, M., Desmond, P., Bridle, N., Tierney, P., Murrie, V., Singh, B., & Copolov, D. (1999). Hippocampal Volume in First-Episode Psychoses and Chronic Schizophrenia: A High-Resolution Magnetic Resonance Imaging Study. *Archives of General Psychiatry*, 56(2), 133–141.
<https://doi.org/10.1001/archpsyc.56.2.133>
- Verduijn, J., Milaneschi, Y., Peyrot, W. J., Hottenga, J. J., Abdellaoui, A., de Geus, E. J. C., Smit, J. H., Breen, G., Lewis, C. M., Boomsma, D. I., Beekman, A. T. F., & Penninx, B. W. J. H. (2017). Using Clinical Characteristics to Identify Which Patients With Major Depressive Disorder Have a Higher Genetic Load for Three Psychiatric Disorders. *Biological Psychiatry*, 81(4), 316–324.
<https://doi.org/10.1016/j.biopsych.2016.05.024>
- Vetlugina, T. P., Lobacheva, O. A., Sergeeva, S. A., Nikitina, V. B., Nevidimova, T. I., & Semke, A. V. (2016). Adjunctive use of interferon γ inducer for treatment of patients

- with schizophrenia. *Acta Neuropsychiatrica*, 28(3), 149–156.
<https://doi.org/10.1017/neu.2015.60>
- Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet Psychiatry*, 3(2), 171–178. [https://doi.org/10.1016/S2215-0366\(15\)00505-2](https://doi.org/10.1016/S2215-0366(15)00505-2)
- Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., Duan, L., Erskine, H., Feigin, V. L., Ferrari, A. J., Fitzmaurice, C., Fleming, T., Graetz, N., Guinovart, C., Haagsma, J., ... Murray, C. J. (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. *The Lancet*, 386(9995), 743–800. [https://doi.org/10.1016/S0140-6736\(15\)60692-4](https://doi.org/10.1016/S0140-6736(15)60692-4)
- Watson, A. J., Giordano, A., Suckling, J., Barnes, T. R. E., Husain, N., Jones, P. B., Krynicki, C. R., Lawrie, S. M., Lewis, S., Nikkheslat, N., Pariante, C. M., Upthegrove, R., Deakin, B., Dazzan, P., & Joyce, E. M. (2022). Cognitive function in early-phase schizophrenia-spectrum disorder: IQ subtypes, brain volume and immune markers. *Psychological Medicine*, 1–10. <https://doi.org/10.1017/S0033291721004815>
- Weinberger, D. R. (1987). Implications of normal brain development for the pathogenesis of schizophrenia. *Archives of General Psychiatry*, 44(7), 660–669.
<https://doi.org/10.1001/archpsyc.1987.01800190080012>
- Wheeler, A. L., & Voineskos, A. N. (2014). A review of structural neuroimaging in schizophrenia: From connectivity to connectomics. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00653>

WHO | *Depression and Other Common Mental Disorders*. (2017). WHO.

http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/

Williams, D. R. (n.d.). *Measuring Discrimination Resource*. 24.

Williams, J., Burgess, S., Suckling, J., Lalousis, P. A., Gkoutos, G., Wood, S., Griffiths, L., Palmer, E., Barsky, A., Karwath, A., David, A. S., Donohoe, G., Neill, J., Deakin, B., Khandaker, G., & Upthegrove, R. (2022). Inflammation and brain structure in schizophrenia and other neuropsychiatric disorders: An interrogation of causal pathways and transcriptomic profiles: Inflammation and brain structure: a Mendelian randomization study with relevance for neuropsychiatric disorders. *JAMA Psychiatry*.

Wise, T., Radua, J., Nartjes, 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.
<https://doi.org/10.1016/j.biopsych.2015.03.004>

Wium-Andersen, M. K., Ørsted, D. D., & Nordestgaard, B. G. (2014). Elevated C-reactive protein associated with late- and very-late-onset schizophrenia in the general population: A prospective study. *Schizophrenia Bulletin*, 40(5), 1117–1127.
<https://doi.org/10.1093/schbul/sbt120>

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
[https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)

Wong, E. H. F., Yocca, F., Smith, M. A., & Lee, C.-M. (2010). Challenges and opportunities for drug discovery in psychiatric disorders: The drug hunters' perspective. *International Journal of Neuropsychopharmacology*, 13(9), 1269–1284. Scopus.
<https://doi.org/10.1017/S1461145710000866>

- Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. F. M., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. T. F., Bigdeli, T. B., Binder, E. B., Blackwood, D. H. R., Bryois, J., Buttenschön, H. N., Bybjerg-Grauholm, J., ... Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, 50(5), 668–681. <https://doi.org/10.1038/s41588-018-0090-3>
- Wu, D., Lv, P., Li, F., Zhang, W., Fu, G., Dai, J., Hu, N., Liu, J., Xiao, Y., Li, S., Shah, C., Tao, B., Zhao, Y., Gong, Q., & Lui, S. (2019). Association of peripheral cytokine levels with cerebral structural abnormalities in schizophrenia. *Brain Research*, 1724, 146463. <https://doi.org/10.1016/j.brainres.2019.146463>
- Yue, Y., Kong, L., Wang, J., Li, C., Tan, L., Su, H., & Xu, Y. (2016). Regional Abnormality of Grey Matter in Schizophrenia: Effect from the Illness or Treatment? *PLOS ONE*, 11(1), e0147204. <https://doi.org/10.1371/journal.pone.0147204>
- Yung, A. R., Yuen, H. P., McGorry, P. D., Phillips, L. J., Kelly, D., Dell’Olio, M., Francey, S. M., Cosgrave, E. M., Killackey, E., Stanford, C., Godfrey, K., & Buckby, J. (2005). Mapping the onset of psychosis: The Comprehensive Assessment of At-Risk Mental States. *The Australian and New Zealand Journal of Psychiatry*, 39(11–12), 964–971. <https://doi.org/10.1080/j.1440-1614.2005.01714.x>
- Zeng, L.-L., Liu, L., Liu, Y., Shen, H., Li, Y., & Hu, D. (2012). Antidepressant Treatment Normalizes White Matter Volume in Patients with Major Depression. *PLOS ONE*, 7(8), e44248. <https://doi.org/10.1371/journal.pone.0044248>
- Zhang, B., Lin, P., Shi, H., Öngür, D., Auerbach, R. P., Wang, X., Yao, S., & Wang, X. (2016). Mapping anhedonia-specific dysfunction in a transdiagnostic approach: An

ALE meta-analysis. *Brain Imaging and Behavior*, 10(3), 920–939.

<https://doi.org/10.1007/s11682-015-9457-6>

Zhang, X., Zhang, Y., Liao, J., Jiang, S., Yan, J., Yue, W., Zhang, D., & Yan, H. (2018). Progressive Grey Matter Volume Changes in Patients with Schizophrenia over 6 Weeks of Antipsychotic Treatment and Their Relationship to Clinical Improvement. *Neuroscience Bulletin*, 34(5), 816–826. <https://doi.org/10.1007/s12264-018-0234-6>

Zhang, Y., Catts, V. S., Sheedy, D., McCrossin, T., Kril, J. J., & Shannon Weickert, C. (2016a). Cortical grey matter volume reduction in people with schizophrenia is associated with neuro-inflammation. *Translational Psychiatry*, 6(12), e982–e982. <https://doi.org/10.1038/tp.2016.238>

Zhang, Y., Catts, V. S., Sheedy, D., McCrossin, T., Kril, J. J., & Shannon Weickert, C. (2016b). Cortical grey matter volume reduction in people with schizophrenia is associated with neuro-inflammation. *Translational Psychiatry*, 6(12), e982. <https://doi.org/10.1038/tp.2016.238>

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 52(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2

Appendix A

Transdiagnostic Structural Neuroimaging features in Depression and Psychosis: A Systematic Review

Paris Alexandros Lalousis, MSc, Stephen Wood, PhD, Renater Reniers, PhD, Lianne Schmal, PhD, Hannah Azam, MSc, Antonella Mazziota, MSc, Hasson Saeed, Msc, Charlotte Wragg, MSc, Rachel Upthegrove, MBBS FRCPsych, PhD

Supplementary Methods and Results

1.1 Search Strategy

For example, PubMed was search with ((“Structural Neuroimaging”[Title/Abstract] OR “Structural MRI”[Title/Abstract] OR sMRI[Title/Abstract] OR MRI[Title/Abstract] OR “Gray Matter Volume”[Title/Abstract] OR "Grey Matter Volume"[Title/Abstract] OR GMV[Title/Abstract] OR “Cortical Thickness”[Title/Abstract] OR “Cortical Folding”[Title/Abstract] OR “Gyrification”[Title/Abstract] OR “Surface Area”[Title/Abstract] OR “White Matter Volume”[Title/Abstract] OR WMV[Title/Abstract] OR “Fractional Anisotropy”[Title/Abstract] OR FA[Title/Abstract] OR “Diffusion Tensor Imaging”[Title/Abstract] OR DTI[Title/Abstract] OR “Diffusion-Weighted Magnetic Resonance Imaging”[Title/Abstract] OR DWI[Title/Abstract] OR “DW-MRI”[Title/Abstract] OR “White Matter Tractography”[Title/Abstract] OR WMT[Title/Abstract])) AND (((Depression[Title/Abstract] OR “Major Depressive Disorder”[Title/Abstract] OR MDD[Title/Abstract] OR “Major Depression”[Title/Abstract] OR MD[Title/Abstract] OR “Depressive Episode”[Title/Abstract] OR “Major Depressive Episode”[Title/Abstract] OR MDE[Title/Abstract] OR “Recent Onset Depression”[Title/Abstract] OR ROD[Title/Abstract] OR “Affective Disorder”[Title/Abstract] OR “Mood Disorder”[Title/Abstract])) AND (Psychosis[Title/Abstract] OR “First Episode Psychosis”[Title/Abstract] OR FEP[Title/Abstract] OR “Recent Onset Psychosis”[Title/Abstract] OR ROP[Title/Abstract] OR Schizophrenia[Title/Abstract] OR “First Episode Schizophrenia”[Title/Abstract] OR FES[Title/Abstract] OR “Schizophrenia Spectrum Disorder”[Title/Abstract] OR SSD[Title/Abstract] OR “Non-Affective Psychosis”[Title/Abstract] OR “Schizophreniform Disorder”[Title/Abstract])).

Study	Selection				Comparability	Exposure			Total
	Adequacy of case definition	Representation of cases	selection of controls	Definition of controls		Ascertainment of exposure	same ascertainment method	Non-response rate	
Coryell et al (2005)	☆				☆☆	☆			4 stars
Cui et al (2020)	☆		☆	☆		☆			4 stars
Jiang et al (2019)	☆	☆	☆	☆	☆	☆			6 stars
Keller et al (2008)	☆	☆	☆	☆	☆	☆	☆	☆	8 stars
Kochunov et al (2013)		☆		☆	☆			☆	4 stars
Meisenzahl et al (2010)		☆		☆	☆☆	☆	☆		6 stars
Ota et al (2013)	☆	☆			☆	☆	☆		5 stars

Table S1: Quality check table.

Study	Analysis	Measured variables	Whole brain analysis done (Y/N)	Results
Coryell et al (2005)	Cortical area-depth and volume	GMV	Yes	Anterior SGPFC did not significantly differ between MDD with psychosis and MDD without psychosis. PsychoticMDD presented smaller volumes in the left-sided posterior SGPFC when comparison between MDD and SCZ post hoc pairwise test was conducted.
Cui et al (2020)	Voxel-wise FA	WMV	Yes	Reduced FA in MDD and SCZ (genu, body and splenium of CC, posterior anterior and superior CR. FA decreases found in genu and body of CC, bilateral anterior and posterior CR, and right superior CR
Jiang et al (2019)	VBM	GMV	Yes	Lower GMV in bilateral MFC in SCZ compared to MDD. Decrease in GMV in MPFC in SCZ than MDD
Keller et al (2008)	ROI analysis	GMV and WMV	No (region of interest analysis) amygdala and Hippocampus	Smaller R and L amygdala volume in MDD with psychosis compared to MDD without psychosis. No significant differences identified by subgroup analysis. No significant hippocampal volume differences.
Kochunov et al (2013)	DTI-FA	WMV	Yes	Age-related WM decline higher in SCZ patients in five specific regions. No significant WM decline in MDD patients.
Meisenzahl et al (2009)	Hippocampal volume	GMV and WMV	Yes	Significance in the volume of the left and right Anterolateral region of the thalamus, ant. medial of the thalamus as well as the Central lateral. A significance as also seen in the FA of the R Subcallosal ACC and L rostral ACC
Ota et al (2013)	ACC, CC, Thalamus, Insula	GMV	ROI	Significant effect of diagnosis on bilateral hippocampal volumes GM and WM

Table S2: Table of results. ROI: region of interest; FA: Fractional Anisotropy; SCZ: schizophrenia; MDD: major depressive disorder; GMV: grey matter volume; WMV: white matter volume; ACC: anterior cingulate cortex; CC: corpus callosum; VBM: voxel-based morphometry; DTI: diffusion tensor imaging; SGPFC: subgenual prefrontal cortex; MPFC: medial prefrontal cortex; R: Right; L: Left.

Appendix B

Transdiagnostic Features, Comorbidity and Classification of Recent Onset Psychosis and Depression: a Multimodal Machine Learning Approach

Paris Alexandros Lalousis, MSc^{1,2}; Stephen J. Wood, PhD^{1,3,4}; Lianne Schmaal, PhD^{3,4}; Katharine Chisholm, PhD^{1,5}; Sian Lowri Griffiths, PhD^{1,2}; Renate L.E.P Reniers, PhD^{1,2,6}; Alessandro Bertolino, MD⁷; Stefan Borgwardt, MD⁸; Paolo Brambilla, MD^{9,10}; Joseph Kambeitz, MD¹¹; Rebekka Lencer, MD, PhD¹²; Christos Pantelis, MB BS, MD, MRCPsych, FRANZCP¹³; Stephan Ruhrmann, MD¹⁴; Raimo K. R. Salokangas, MD, PhD, MSc¹⁵; Frauke Schultze-Lutter, PhD¹⁶; Carolina Bonivento, PhD¹⁷; Dominic Dwyer, PhD¹¹; Adele Ferro, PsyD, PhD¹⁰; Theresa Haidl, MD¹⁴; Marlene Rosen, MSc¹⁴; Andre Schmidt, PhD⁸; Eva Meisenzahl, MD¹⁶; Nikolaos Koutsouleris, MD^{11*}; Rachel Upthegrove, MBBS FRCPsych, PhD^{1,2,18*} and the PRONIA Consortium#

* Joint Senior Authors

#See PRONIA consortium author list

Author Affiliations:

- 1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom
- 2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom
- 3 -Orygen, the National Centre of Excellence in Youth Mental Health; Melbourne, Australia
- 4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia
- 5 -Department of Psychology, Aston University, United Kingdom
- 6 –Institute of Clinical Sciences, University of Birmingham, United Kingdom
- 7 -Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy
- 8 -Department of Psychiatry, University of Basel, Basel, Switzerland
- 9 -Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy
- 10 - Department of Neurosciences and Mental Health, IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.
- 11 -Department of Psychiatry and Psychotherapy, Ludwig Maxmilians University, Munich, Germany
- 12 -Department of Psychiatry, University of Münster, Münster, Germany
- 13 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia
- 14 -Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany
- 15 -Department of Psychiatry, University of Turku, Turku, Finland

16 -Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

17 -IRCCS “E. Medea” Scientific Institute, San Vito al Tagliamento (Pn), Italy

18 -Early Intervention Service, Birmingham Women’s and Children’s NHS trust, United Kingdom

Keywords: psychosis, depression, transdiagnostic, machine learning, MRI, comorbidity

Corresponding Author:

Mr. Paris Alexandros Lalousis, BSc, MSc

School of Psychology

University of Birmingham, 52 Pritchatts Road

B15 2SA, Birmingham, United Kingdom

Supplementary Methods and Results

1. Supplementary Methods

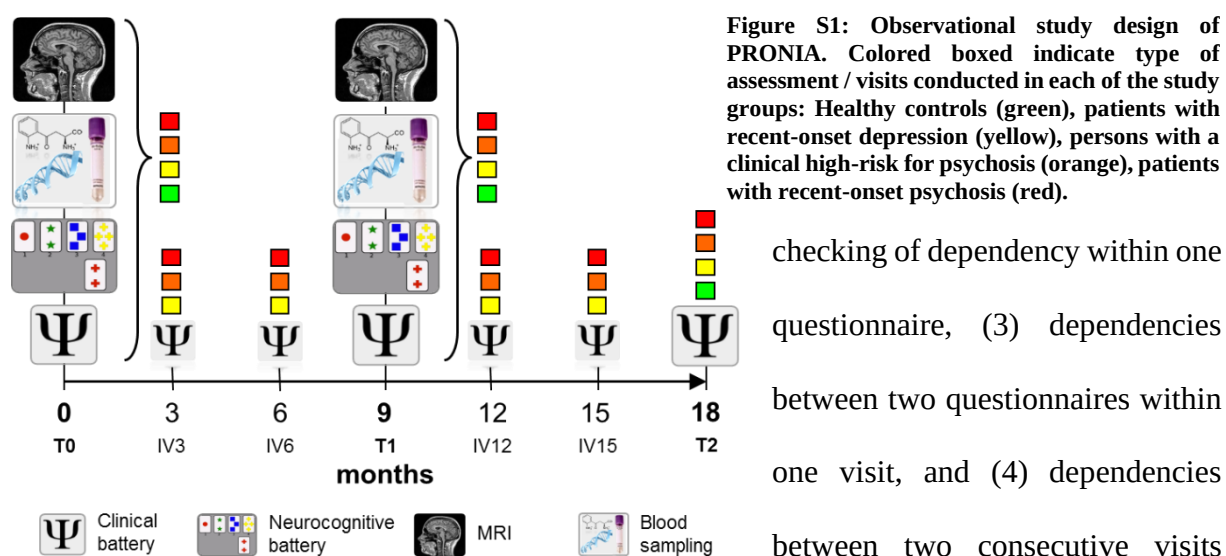
PRONIA recruitment infrastructure

The 300 study participants (154 individuals with ROP and 146 individuals with ROD) analyzed in the present study were recruited following a standardized recruitment and ascertainment protocol (see **Figure S1** and **Table S3**). The observational study protocol involved follow-up examinations every three months after the index ascertainment and was implemented by the following 7 PRONIA sites:

Table S1: Characteristics of the recruiting institutions in the PRONIA consortium.

PRONIA Site	Institution Name	Country	Type of Service	Catchment Population	Screening population / year
Munich	Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich	DE	Academic outpatient services including specialized service for early recognition of psychosis; tertiary care academic hospital	1,200,000	700
Basel	Department of Psychiatry and Psychotherapy, University of Basel	CH	Academic inpatient and outpatient services including specialized service for early recognition and intervention of psychosis; tertiary care academic hospital	500,000	200
Milan Niguarda	Department of Pathophysiology and Transplantation, University of Milan. Four recruitment hospitals: Niguarda, Policlinico, San Paolo, Villa San Benedetto Menni in Albese con Cassano	IT	Psychiatric outpatient services including specialized services for early recognition of psychosis and persons at high risk; Academic hospital, providing psychiatric inpatient services, psychiatric outpatient services and local services;	600,000	1,000
Cologne	Department of Psychiatry and Psychotherapy, University of Cologne	DE	Academic outpatient services including specialized service for early recognition of psychosis; tertiary care academic hospital	1,000,000	600
Birmingham	The University of Birmingham	UK	Academic specialised Early Intervention Service for Psychosis covering Birmingham and Solihull. Community and Inpatient	1,200,000	800
Turku	Department of Psychiatry, University of Turku	FI	Psychiatric outpatient and hospital services responsible for treatment of psychiatric patients in their catchment areas in the South-Western Finland.	284,000	2,300
Udine	Department of Psychiatry, University of Udine	IT	Psychiatric outpatient services, academic hospital and local services. Tertiary care neuropsychiatric service	600,000	500

Upon study enrolment, the participants were pseudonymized twice, locally at each site and centrally at the level of the PRONIA portal. The PRONIA portal consists of a multi-user database hosting the clinical and neurocognitive information, and defaced MR images obtained from the study participants. The data are organized into digital questionnaires, visits, and cases. The portal provides the case managers with a controlled web-based interface to enter and upload the different data into the respective questionnaires. Furthermore, the PRONIA consortium has implemented a PRONIA@home mobile device interface that allows the study participants to securely log into the portal and fill out the self-rating questionnaires of given visit. Upon completion of the data entry across all questionnaires of given visit, the data are checked by an automatic quality control procedure which executes approximately 1600 data integrity and dependency rules. These rules include (1) basic checking of missing data and data ranges, (2)



(such as consistency of dates). Detected errors are fed back to the respective case managers allowing for a manual correction of the respective issues. This process is re-iterated until the quality of the clinical questionnaires in the given visit is sufficient for the entire visit to be locked.

PRONIA study design and examination instruments

A comprehensive battery of ascertainment tools was used within a longitudinal observational study design to generate a multi-modal phenotypic profile of each study participant (see **Figure**

S1 and Table S2). The clinical part of the battery compiled questionnaires that capture sociodemographic, somatic, environmental, diagnostic, psychopathological, functional and quality-of-life related variables in the PRONIA study population. This battery was complemented by multi-domain neurocognitive and neuroimaging examinations as well as blood sampling for later genetic characterization, which were carried out at the baseline and 9-month follow-up timepoints.

Table S2: Clinical and neurocognitive examinations performed in the CHR, ROD, ROP, and HC groups during the 18-month follow-up period of the study. Clinical assessment types:

OR Observer-based rating instrument, SR Self-rating-based instrument. **Examination timepoints:** T0 Baseline examination, IV3/IV6/IV12/IV15 3, 6, 12, 15-month examinations conducted only in the clinical study participants, T1 9-month examination, T2 18-month follow-up examination. **Observer-based instruments:** CAARMS Comprehensive Assessment of the At-Risk Mental States(Yung et al., 2005), CHR Criteria Clinical High-Risk criteria summary questionnaire, FROGS Functional Remission in General Schizophrenia(Llorca et al., 2009), GAF Global Assessment of Functioning, GF:S/R Global Functioning: Social / Role(Cornblatt et al., 2007), PANSS Positive and Negative Symptom Scale(Kay et al., 1987), PAS Premorbid Adjustment Scale(Alvarez et al., 1987), SANS Scale for the Assessment of Negative Symptoms(Andreasen, 1989), SCID-IV Screening/Summary Structured Clinical Interview for DSM-IV(First, Michael B et al., 2002), SIPS Standardized Interview for the Assessment of Prodromal Symptoms (modified version 5.0)(McGlashan et al., 2001), SPI-A [COGDIS/COPER] Schizophrenia Proneness Instrument [Cognitive disturbances (COGDIS) / Cognitive-Perceptual (COPER) disturbances](F. Schultze-Lutter et al., 2007), Transition Criteria Interval questionnaire for the assessment of transition criteria, UHR - Schizotypy, Genetic Risk Interview for the Assessment of Schizotypal personality traits, and familial risk for psychosis. **Self-rating instruments:** BDI-II Beck Depression Inventory II(A. Beck et al., 1996), CISS-24 Coping Inventory for Stressful Situations – 24 items(Endler, 2004), CTQ Childhood Trauma Questionnaire(Bernstein et al., 1994), EHI-SR Edinburgh Handedness Inventory – Short Version(Veale, 2014), EDS Everyday Discrimination Scale – Modified Version(D. R. Williams, n.d.), LEE Level of Expressed Emotions(Cole & Kazarian, 1988), MSPSS the Multidimensional Scale for Perceived Social Support(Zimet et al., 1988), NEO-FFI NEO Five Factor Inventory of Personality Traits(Costa et al., 1992), RSA Resilience Scale for Adults(Friborg et al., 2003), SPIN Social Phobia Inventory(Connor et al., 2000), WHO-QOL-BREF WHO Quality of Life Questionnaire-Brief Version(Health, 1996). Neurocognitive tests: CPT-IP Continuous-Performance Test-Identical Pairs (adapted tablet version)(Cornblatt et al., 1988), DANVA Diagnostic Analysis of Non-Verbal Accuracy 2 (adapted tablet version)(Nowicki & Duke, 2001), DS Auditory Digit Span (Forward/Backward) adapted from the PEBL battery, DSST Digit-Symbol-Substitution Test from the BACS battery, ROCF Rey-Osterrieth complex figure(Rey, 1941), SAT Salience Attribution Task (adapted version)(Roiser et al., 2010), SOPT self-ordered pointing task (adapted version)(Petrides & Milner, 1982), TMT-A/-B Trail-Making Test A and B(Reitan, 1958), VF phonemic/semantic verbal fluency test.

Instrument	Form	Screening		T0		IV3	IV6	T1		IV12	IV15	T2	
		PAT	HC	PAT	HC	PAT	PAT	PAT	HC	PAT	PAT	PAT	HC
General Data	OR	X	X					X	X			X	X
Reasons for Referral	OR	X											
Treatment Documentation	OR	X	X			X	X	X	X	X	X	X	X
Somatic state and Health History	OR	X	X					X	X			X	X
SPI-A COGDIS/COPER	OR	X	X			X	X	X	X	X	X	X	X
SIPS positive symptoms	OR	X	X			X		X	X			X	X
CAARMS	OR	X	X			X		X	X			X	X
GAF	OR	X	X			X		X	X			X	X
UHR – Schizotypy, Genetic Risk	OR	X	X			X		X	X			X	X
CHR Criteria	OR	X	X					X	X			X	X
Transition Criteria	OR					X	X			X	X		
SCID-IV Screening	OR	X	X					X	X			X	X
SCID-IV Summary	OR	X	X					X	X			X	X
Demographic and Biographic Data	OR			X	X			X	X			X	X

PAS	OR			X	X			X				X	
SPI-A	OR			X	X			X				X	
SIPS negative, disorganized and general symptoms	OR			X	X			X				X	
PANSS	OR			X		X	X	X		X	X	X	
SANS	OR			X				X				X	
Chart of Life Events	OR			X	X	X	X	X	X	X	X	X	X
FROGS	OR			X				X				X	
GF: Social & Role	OR			X	X	X	X	X	X	X	X	X	X
Prognostic evaluation	OR			X				X				X	
MSPSS	SR			X	X			X	X			X	X
RSA	SR			X	X			X	X			X	X
CISS 24	SR			X	X			X	X			X	X
SPIN	SR			X	X			X	X			X	X
BDI-II	SR			X	X	X	X	X	X	X	X	X	X
WHO-QOL-BREF	SR			X	X			X	X			X	X
EHI-SR	SR			X	X								
LEE	SR			X	X			X	X			X	X
Wisconsin Scales	SR			X	X								
EDS	SR			X	X								
Bullying Scale T0	SR			X	X								
CTQ	SR			X	X								
NEO-FFI	SR			X	X								
DS backward (BACS)	NPT			X	X			X	X				
DS forward (BACS)	NPT			X	X			X	X				
CPT-IP (BACS)	NPT			X	X			X	X				
DANVA	NPT			X	X			X	X				
DSST	NPT			X	X			X	X				
RAVLT	NPT			X	X			X	X				
ROCF	NPT			X	X			X	X				
SAT	NPT			X	X			X	X				
SOPT	NPT			X	X			X	X				
TMT-A	NPT			X	X			X	X				
TMT-B	NPT			X	X			X	X				
VF phonetic	NPT			X	X			X	X				
VF semantic	NPT			X	X			X	X				
WAIS-III	NPT			X	X			X	X				

MRI harmonization and data acquisition

When setting up the PRONIA study, we decided to generate a MRI database that would represent the MR scanner sequence heterogeneity encountered in clinical real-world. The aim of this strategy was to strengthen the generalizability and clinical applicability of the predictive models developed by our machine learning analyses. Thus, we agreed on a minimal harmonization protocol that required the PRONIA sites to only (1) acquire isotropic or nearly isotropic voxel sizes of preferably 1 mm resolution, (2) set the Field Of View (FOV) parameters accordingly to guarantee the full 3D coverage of the brain including all parts of the cerebellum, and (3) define the relaxation time (TR) and echo time (TE) as well as other imaging parameters in a way that would maximize the contrast between cortical ribbon and the white matter and enhance the signal-to-noise ratio in the images. At every site all the images were visually inspected, automatically defaced and anonymized using an in-house Freesurfer-based script

before the data was centralized. **Table S3** lists the parameters defining the structural MR sequences used to examine in the PRONIA discovery sample participants.

Table S3: MR scanner systems and structural MRI sequence parameters used at the respective PRONIA sites.

PRONIA Site	Model	Field Strength	Coil Channels	Flip Angle	TR [ms]	TE [ms]	Voxel Size [mm]	FOV	Slice Number
Munich	Philips Ingenia	3T	32	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Milan Niguarda	Philips Achieva Intera	1.5T	8	12	Shortest (8.1)	Shortest (3.7)	0.93 x 0.93 x 1.0	240 x 240	170
Basel	SIEMENS Verio	3T	12	8	2000	3.4	1.0 x 1.0 x 1.0	256 x 256	176
Cologne	Philips Achieva	3T	8	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Birmingham	Philips Achieva	3T	32	8	8.4	3.8	1.0 x 1.0 x 1.0	288 x 288	175
Turku	Philips Ingenuity	3T	32	7	8.1	3.7	1.0 x 1.0 x 1.0	256 x 256	176
Udine	Philips Achieva	3T	8	12	Shortest (8.1)	Shortest (3.7)	0.93 x 0.93 x 1.0	240 x 240	170

MRI processing pipeline

The manual of the CAT12 toolbox (<http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>)

details the processing steps applied to the structural images. These steps consist of:

- (1) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering(Manjón et al., 2008).
- (2) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local variations of intensity distributions as slowly varying spatial functions and thus achieves a homogeneous segmentation across cortical and subcortical structures(Rajapakse et al., 1997).
- (3) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior information of adjacent voxels into the segmentation estimation generated by AMAP(Rajapakse et al., 1997).
- (4) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM) inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP segmentation.

- (5) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the AMAP-generated tissue segments.
- (6) A high-dimensional DARTEL registration of the image to a MNI-template generated from the MRI data of 555 healthy controls in the IXI database (<http://www.braindevelopment.org>).
- (7) The GM maps were then multiplied with the Jacobian determinants that were obtained during registration in order to produce GM volume maps.
- (8) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

1.5. Support Vector Machine Learning Preprocessing

Prior to the machine learning analysis the following preprocessing steps were completed: a) every feature was scaled from 0 to 1 and completely non-finite features were zeroed-out; b) any feature that had infinite values was pruned; c) missing values were imputed using KNN Euclidean distance median replacement using 7 nearest neighbours. For each missing value of a given CV1 or CV2 subject, a subset of cases that had values for the given variable and had values in all other variables which were non-empty were identified. Subjects in the source subset were sorted according to their similarity with the target subject using the Euclidean distance. Then, the median of the given variable was computed using the 7 nearest neighbours. The original, non-imputed training matrix was used at all times d) site effects were corrected for. This step consisted of i) performing a principal components analysis, ii) identifying the components that are most correlated with site variance, and iii) reconstructing the data without these components in order to effectively remove such effects from the data. Any component with a Spearman correlation coefficient greater than 0.5 was excluded from the analysis and 90% of the variance was retained. e) the data was scaled again. The model

performance criterion that was used was balanced accuracy and the learning algorithm that was chosen was LIBSVM. Imbalanced learning was corrected for by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes (weighting the hyperplane). The kernel type that was linear with eleven learning parameters in order to optimize the choice of C value. Wrapper methods were activated at all parameter combinations with greedy sequential backward feature selection (Stop at k=90% of features; Feature stepping at 10% of worst performing features at each cycle). Starting with the full feature set the SVM was ran iteratively with the 10% worst performing features being eliminated at each cycle until 10% of the feature pool was left. In order to assess the models' statistical significance we performed permutation analysis to create an empirical null distribution of weights for each feature and then compare the observed weight to this distribution. The models were retrained in the cross-validation framework using the respective feature and label subsets obtained from the observed-label analyses 1000 times. For each permutation the predictions were accumulated into a permuted ensemble prediction for each CV2 subject. In that way a null distribution of out-of-training classification performance for the prediction models was produced. The significance of the observed out-of-training classification performance was calculated as the number of events where the permuted out-of-training classification performance was higher or equal to the observed classification performance divided by the number of permutations performed. Then the significance of the model was determined according to a p threshold of $p < 0.05$.

For the neuroimaging model the same steps used apart from: 1) when inputting the grey matter volume (GMV) images into the modality a G-theory mask (Mushquash & O'Connor, 2006) was used for quantifying the degree of reliability for the imaging modalities based on a travelling participants study and 2) during preprocessing of the neuroimaging model the following additional steps were completed a) any feature that had infinite values was pruned;

b) features were ranked up using external ranking by means of hard selection thresholds (15%, 25%, and 50%). Only the top 15%, 25%, and 50% features as defined by the G-theory mask were kept; c) Principal Component Analysis (PCA) Dimensionality Reduction (dimensions: 0.8); d) every feature was scaled from 0 to 1 and completely non-finite features were zeroed out; 3) when visualizing the classification a 8 FWHM kernel was used to smooth the images.

Finally, a stacking-based data fusion framework(R. Polikar, 2006; Wolpert, 1992) was used to examine whether the combination of the clinical-based and the neuroimaging-based models would provide a superior classification accuracy. To achieve this, the decision scores of the neuroimaging and the clinical models were combined, standardized, and forwarded to a greedy sequential forward search algorithm(Saeys et al., 2007) which found a parsimonious combination of classifiers maximizing $\overline{\text{PSI}}_{\text{reg}}$ across the C parameter range by employing L2-regularized logistic regression (L2LR)(Fan et al., 2008). Each L2LR ensemble was then applied to the standardized neuroimaging and clinical based decision scores available for the CV₂ validation data. Class prediction was achieved by using majority voting on Majority voting on P_{ens} .

1.6. Hyperparameter Optimization

For hyperparameter optimization, we computed

$$\overline{\text{BAC}}_{\text{reg}} = \sum_{i=1}^k \left(\frac{n_{TP_i}}{n_{TP_i} + n_{FN_i}} + \frac{n_{TN_i}}{n_{TN_i} + n_{FP_i}} \right) / 2 \text{ at given parameter combination across all } k \text{ CV}_1$$

partitions with $n_{TP_i}/n_{TP_i} + n_{FN_i}$ being Sensitivity and $n_{TN_i}/n_{TN_i} + n_{FP_i}$ being Specificity, and the fraction of the training population serving as support vectors in the i^{th} CV₁ partition. Our optimization technique's aim was to find a combination of T_G , PC s and the SVM's regularization parameter C [range: $2^{[-3 \rightarrow +4]}$] that maximized $\overline{\text{BAC}}_{\text{reg}}$ within a 3 (T_G) $\times$ 5 (PC) $\times$ 8 (C) hyperparameter cube. The optimized ensemble was then applied to the CV₂ validation

data to produce a mean decision score ($\overline{D_{ens}}$) and majority voting-based class membership probabilities (P_{ens}) for each CV₂ validation subject. This produced a mean decision score ($\overline{(D_{ens})}$) and majority voting-based class membership probabilities (P_{ens}) for each CV₂ validation subject.

1.7. Stacking Fusion Framework

The decision scores of the neuroimaging and the clinical models were combined, standardized, and forwarded to a greedy sequential forward search algorithm(Saeys et al., 2007) which found a parsimonious combination of classifiers maximizing $\overline{PSI_{reg}}$ across the C parameter range by employing L2-regularized logistic regression (L2LR)(Fan et al., 2008). Each L2LR ensemble was then applied to the standardized neuroimaging and clinical based decision scores available for the CV₂ validation data. Class prediction was achieved by using majority voting on Majority voting on P_{ens} .

1.8. Clinical/Neurocognitive Model List of Tests

The clinical and neurocognitive tests used to train the clinical support vector machine learning model were the following: (1) Wisconsin Schizotypy Scale (WSS) physical and social anhedonia subscales(Barkus & Badcock, 2019; Cannon-Spoor et al., 1982; Shankman et al., 2010); (2) Premorbid Adjustment Scale (PAS)(Romm et al., 2011; Silverstone & Salsali, 2003), (3) Functional Recovery Scale in Schizophrenia (FROGS)(Llorca et al., 2009); (4) SIPS-negative(Kline et al., 2018); (5) Schizophrenia Proneness Instrument, Adult version (SPI-A)(Schultze-Lutter et al., 2007); (6) Resilience Scale for Adults (RSA)(Friborg et al., 2003; M.-H. Han & Nestler, 2017; Marulanda & Addington, 2016); and a range of neurocognitive tests: (1) Digit Span Test (DST)(McClintock et al., 2010; Sánchez-Torres et al., 2015); (2) Phonemic Verbal Fluency (PVF)(Becker et al., 2010; McClintock et al., 2010); (3) Semantic Verbal Fluency (SVF)(Becker et al., 2010; McClintock et al., 2010); (4)

Auditory-Verbal Learning Test (AVLT)(McClintock et al., 2010), and (5) Digit Symbol Substitution Test (DSST)(McClintock et al., 2010; Sánchez-Torres et al., 2015).

Table S4: DGPPN S3 Guidelines for the treatment of first-episode psychosis and schizophrenia (translated English version of Table 4.1 stated in the short version of the guideline manual available in [https://www.dgppn.de/ Resources/Persistent/a6e04aa47e146de9e159fd2ca1e6987853a055d7/S3_Schizo_Kurzversion.pdf](https://www.dgppn.de/Resources/Persistent/a6e04aa47e146de9e159fd2ca1e6987853a055d7/S3_Schizo_Kurzversion.pdf)). Candidate CHR and ROD patients were excluded if they had received antipsychotic medication (1) for more than 30 cumulative days at or above the minimum target dosage threshold for the treatment of first-episode psychosis, or (2) within the past 3 months before psychopathological baseline assessments at or above the minimum target dosage threshold for the treatment of first-episode psychosis. Abbreviations: DI dosage interval, ²maximum recommended dosage according to prescribing information.

Substance	Recommended starting dosage (mg/d)	DI'	Target dosage first-episode psychosis (mg/d)	Target dosage relapsing schizophrenia (mg/d)	Maximum dosage recommended (mg/d) ²
Atypical Antipsychotics					
Amisulpride	200	(1)-2	100-300	400-800	1200
Aripiprazole	(10)-15	1	15-(30)	15-30	30
Olanzapine	5-10	1	5-15	5-20	20
Quetiapine	50	2	300-600	400-750	750
Risperidone	2	1-2	1-4	3-6-(10)	16
Ziprasidone	40	2	40-80	80-160	160
Typical Antipsychotics					
Fluphenazine	0.4-10	2-3	2.4-10	10-20	20-(40)
Flupentixole	2-10	1-3	2-10	10-60	60
Haloperidole	1-10	(1)-2	1-4	3-15	100
Perazine	50-150	1-2	100-300	200-600	1000
Perphenazine	4-24	1-3	6-36	12-42	56
Pimozide	1-4	2	1-4	2-12	16
Zotepine	25-50	2-(4)	50-150	75-150	450
Zuclopenthixole	2-50	1-3	2-10	25-50	75

Table S5: SCID DSM-IV diagnoses detailed breakdown in ROP groups (Discovery Sample)

	ROP total N=154
DSM Diagnosis	N (%)
Schizophrenia	72 (46.7)
Schizophreniform	13 (8.0)
Psychosis NOS	18 (11.7)
Delusional Disorder	11 (7.1)
Brief psychotic disorder	6 (3.8)
Substance Induced	3 (1.9)
Schizoaffective Disorder	10 (6.5)
Bipolar Disorder	12 (7.8)
Severe depression with psychosis	9 (5.7)

Table S6: Demographics of Replication Sample

	ROP Group	ROD Group	t/χ^2	P Value	Pure ROP Group	Pure ROD Group	t/χ^2	P Value	ROP+D Group	ROD+P Group	t/χ^2	P Value
Sample Sizes, No.	161	131			50	53			90	25		
Age, Mean (SD)	25.6 (6.1)	25.0 (5.9)	$t = .787$	.432	25.8 (6.3)	25.7 (6.1)	$t = .039$	.969	25.3 (6.1)	25.5 (6.5)	$t = -.156$	.876
Sex (Male/Female)	89/71	68/62	$\chi^2 = .339$	.844	30/20	24/29	$\chi^2 = 2.234$	.135	48/42	15/10	$\chi^2 = .351$	.554

2. Supplementary Results

A number of exploratory supplementary analyses are presented in the supplementary results. These include a) correlational analyses between the clinical and imaging decision scores of the four different groups (Pure ROP/Pure ROD/ROP+D/ROD+P), in order to explore whether there was an association between clinical, and neuroanatomical features driving misclassification; b) association and comparison analyses between patients who were correctly classified and patients who were misclassified, to explore any potential ‘sub-group’ identification of misclassified participants; c) whole brain GMV comparison between the whole ROP and the whole ROD group, between subjects who were classified as ROP and subjects who were classified as ROD; and d) group level comparisons of the decisions scores of the three machine learning models.

2.1. Correlation of Clinical and Imaging Decision scores.

We wanted to examine the relationship between the clinical and imaging decision scores of our different groups (Pure ROP/Pure ROD/ROP+D/ROD+P) employing a Pearson correlation test. In figure S2 the non-significant positive correlation between the clinical and imaging decision scores in the Pure ROP group can be seen ($r = 0.005$, $p < .05$). In figure S3 we can see the positive correlation between the clinical and imaging decision scores in the Pure ROD group ($r = .112$, $p = .073$). Figure S4 shows the positive correlation between the clinical and imaging decision scores in the ROP+D group ($r = .062$, $p < .05$). Finally,

figure S5 shows the positive correlation between the clinical and imaging scores in the ROD+P group ($r = .387$, $p = 0.14$).

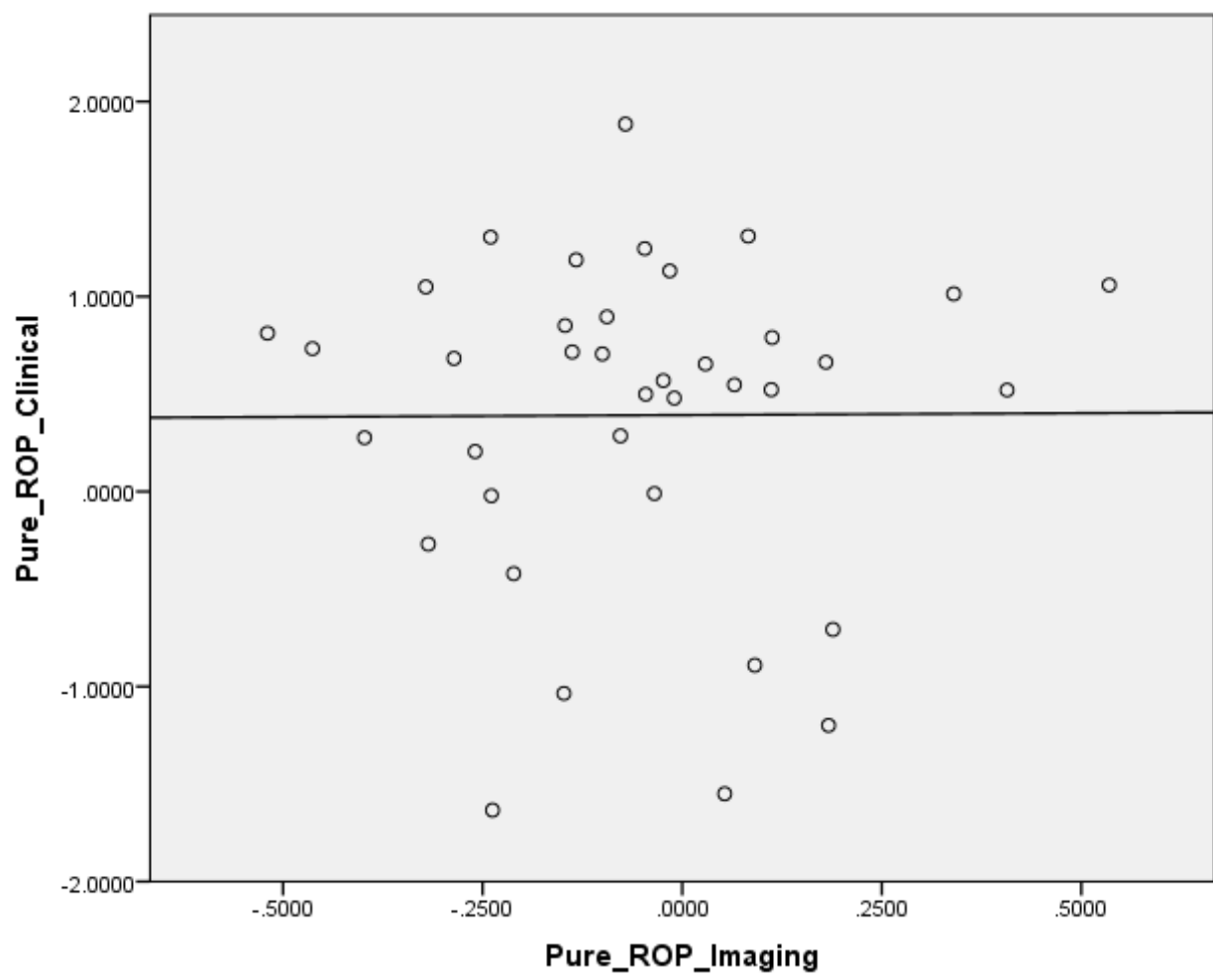


Figure S2. Pure ROP Clinical and Imaging Decision Scores Correlation

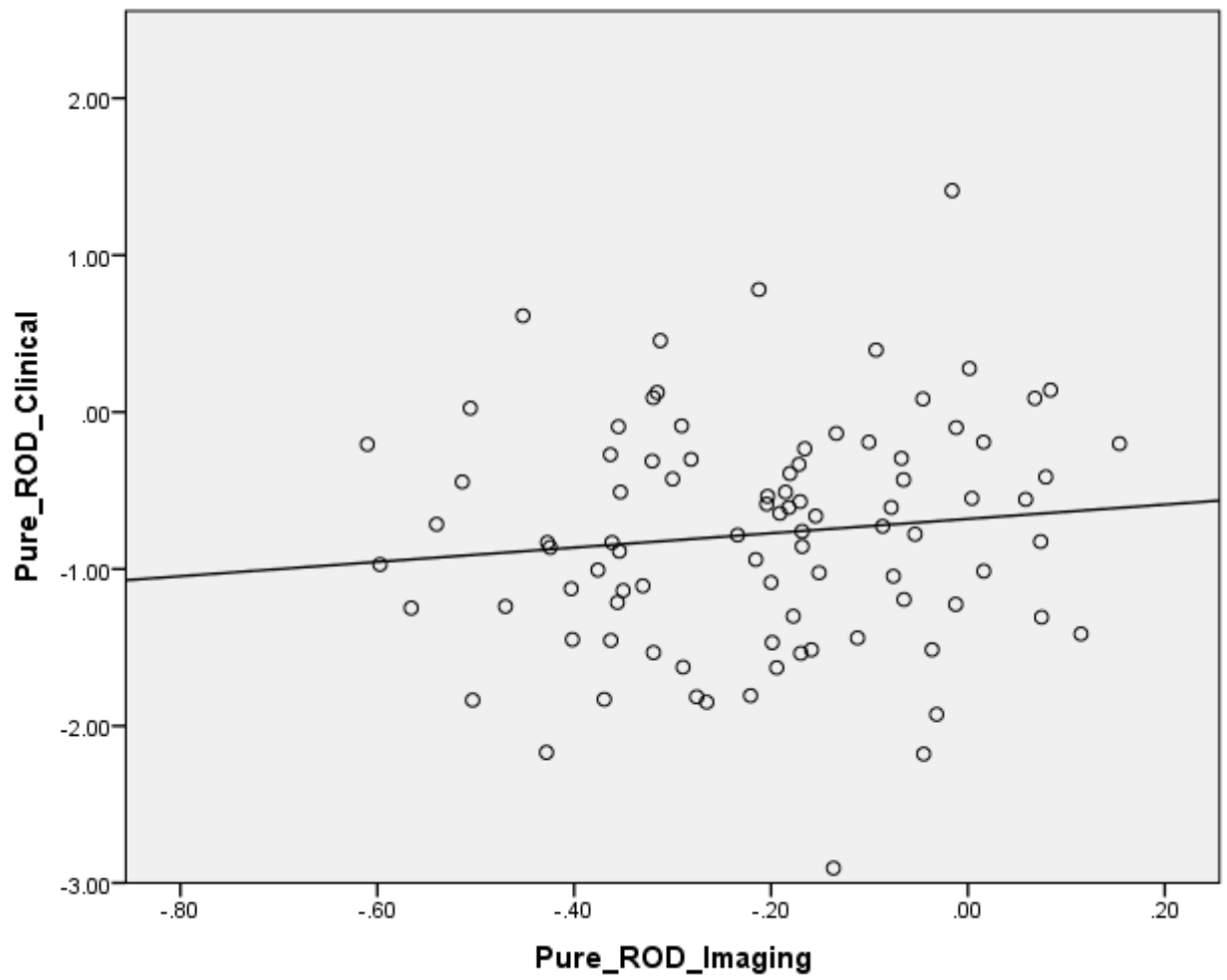


Figure S3. Pure ROD Clinical and Imaging Decision Scores Correlation

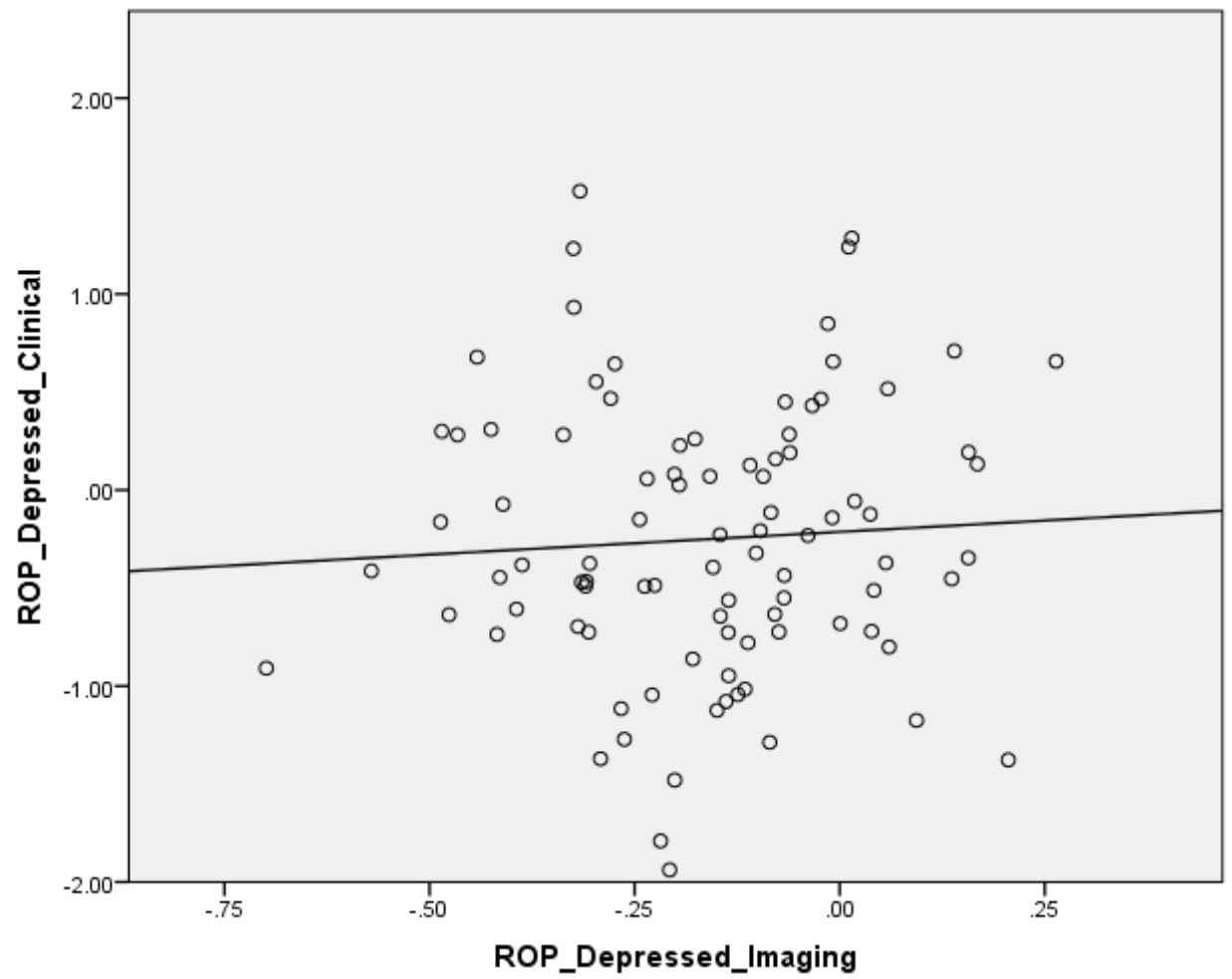


Figure S4. ROP Depressed Clinical and Imaging Decision Scores Correlation

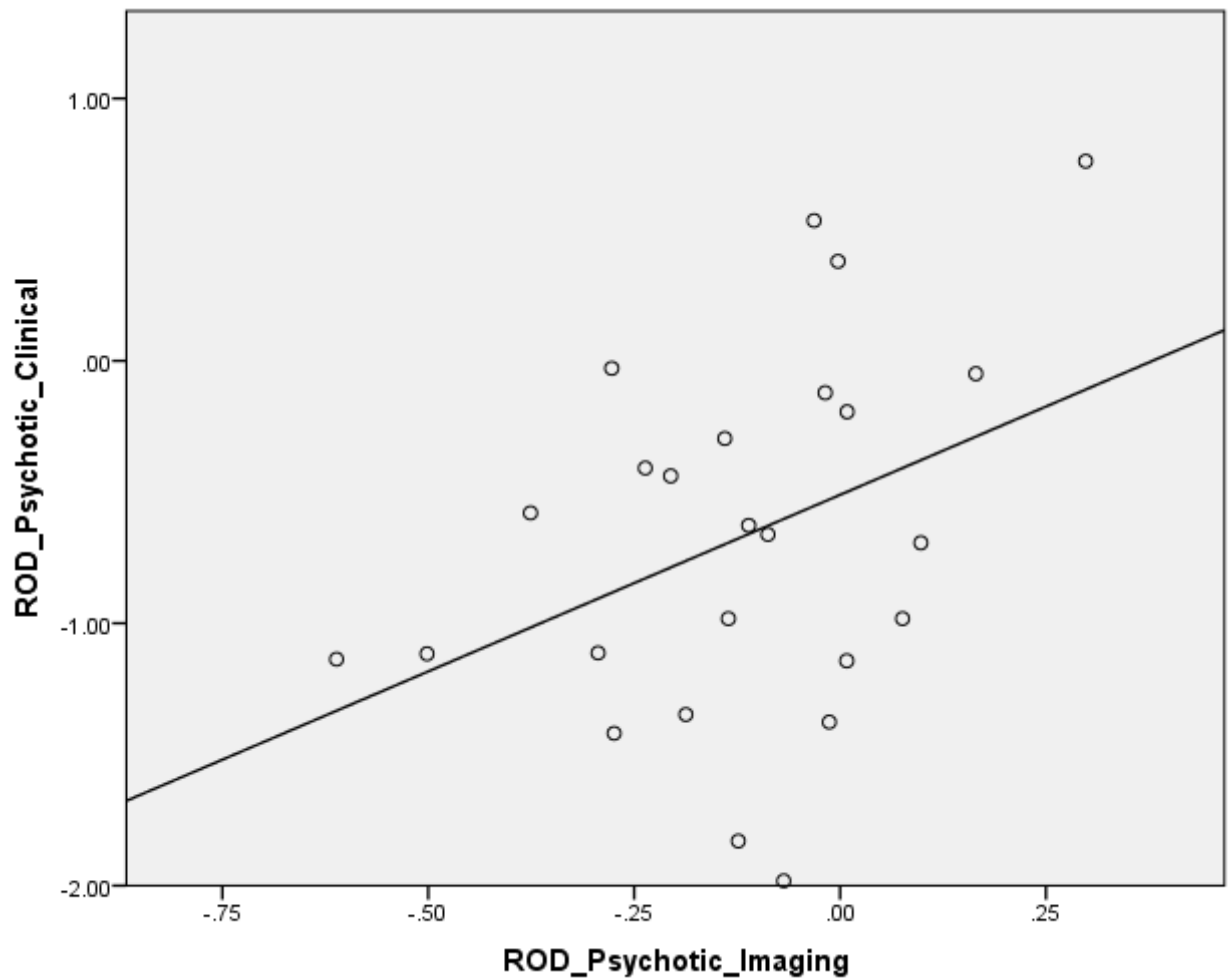


Figure S5. ROP Depressed Clinical and Imaging Decision Scores Correlation

2.2. Association and comparison analysis between patients who were correctly classified and patients who were misclassified

To understand the reasons why some of the patients in our sample were misclassified we performed a series of association and comparison analysis between those patients who were correctly classified and those who were misclassified using clinically relevant measures. In table S5 we can see the association between Pure ROP patients who were correctly classified as ROP and Pure ROP patients who were misclassified as ROD both in the clinical and imaging models. In Table S6 we have compared Pure ROP patients who have been correctly classified and Pure ROP patients who have been misclassified in

seven clinically relevant measures (PANSS Positive, WSS Perceptual Aberration, CAARMS Unusual Thought Content, CAARMS Non-Bizarre Ideas, and CAARMS Disorganized Speech), both in the clinical and the imaging models. Furthermore we performed an association analysis between ROP+D patients who had been correctly classified and ROD+D patients who had been misclassified (Table S7). Finally, correctly classified ROP+D patients were compared to misclassified ROP+D patients in the same seven measures that were mentioned previously (Table S8).

Table S7. Association Analyses between Pure ROP Classification and Schizotypy Criteria

		SPD Criteria Not Met Count	SPD Criteria Not Met Expected Count	SPD Criteria Met Count	SPD Criteria Met Expected Count	χ^2	P Value
Clinical							
Classification	Correct Classification	23	23.8	3	2.2		
	Misclassification	10	9.2	0	0.8	1.25	.131
Imaging							
Classification	Correct Classification	12	12.8	2	1.2		
	Misclassification	21	20.2	1	1.8	1.06	.151

Table S8. Comparison Analyses between Pure ROP Correctly Classified and Pure ROP Misclassified (P-Values with an asterisk survive FDR (Benjamini-Hochberg) correction threshold)

		Correctly Classified Mean and Standard Deviation	Misclassified Mean and Standard Deviation	<i>t</i>	P Value
--	--	---	--	----------	---------

Clinical					
Classification	PANSS Positive Total	16 (5.3)	13.7 (8.0)	1.013	.159
	WSS Magical Ideation				
	Total	2.0 (2.0)	3.5 (2.7)	-1.780	.041*
	WSS Perceptual				
	Aberration Total	.5 (.9)	1.1 (1.8)	-1.208	.117
	CAARMS Unusual				
	Thinking Content	4.8 (2.1)	5.2 (1.9)	-.485	.315
	CAARMS Non Bizarre				
	Ideas	4.6 (2.1)	2.7 (2.9)	2.203	.017*
	CAARMS Perceptual				
	Abnormalities	3.6 (2.7)	5.2 (1.4)	-1.740	.045*
	CAARMS				
	Disorganized Speech	2.4 (2.0)	2.0 (2.4)	.544	.295
Imaging					
Classification	PANSS Positive Total	13.1 (4.5)	16.8 (6.6)	-1.888	.034
	WSS Magical Ideation				
	Total	2.8 (2.4)	2.1 (2.1)	.965	.170
	WSS Perceptual				
	Aberration Total	.4 (1.3)	.8 (1.2)	-.846	.201
	CAARMS Unusual				
	Thinking Content	4.6 (2.1)	5.0 (2.0)	-.599	.276
	CAARMS Non Bizarre				
	Ideas	4.4 (2.3)	3.8 (2.5)	.724	.237
	CAARMS Perceptual				
	Abnormalities	4.6 (2.2)	3.6 (2.6)	1.123	.134
	CAARMS				
	Disorganized Speech	2.2 (1.9)	2.3 (2.2)	-.114	.455

Table S9. Association Analyses between ROP Depressed Classification and Schizotypy Criteria

		SPD Criteria Not Met Count	SPD Criteria Not Met Expected Count	SPD Criteria Met Count	SPD Criteria Met Expected Count	χ^2	P Value
Clinical							
Classification	Correct Classification	28	30.3	6	3.7		
	Misclassification	54	51.7	4	6.3	2.55	.055
Imaging							
Classification	Correct Classification	16	16	2	2		
	Misclassification	66	66	8	9	.001	.485

Table S10. Comparison Analyses between ROP Depressed Correctly Classified and ROP Depressed Misclassified (P-Values with an asterisk survive FDR (Benjamini-Hochberg) correction threshold)

		Correctly Classified Mean and Standard Deviation	Misclassified Mean and Standard Deviation	<i>t</i>	P Value
Clinical					
Classification	PANSS Positive Total	20.0 (7.2)	17.9 (5.3)	1.609	.055*
	WSS Magical Ideation Total	5.1 (3.0)	4.1 (3.1)	1.540	.063
	WSS Perceptual Aberration Total	2.5 (2.7)	3.0 (3.5)	-.749	.228

Imaging Classification	CAARMS Unusual Thinking Content	4.8 (1.9)	5.1 (1.6)	-.651	.258
	CAARMS Disorganized Speech	2.4 (2.2)	1.3 (1.6)	2.497	.007*
	CAARMS Non Bizarre Ideas	5.06 (1.9)	4.6 (2.1)	1.042	.150
	CAARMS Perceptual Abnormalities	3.1 (2.5)	3.4 (2.5)	-.505	.307
	PANSS Positive Total	20.4 (6.3)	18.2 (6.0)	1.495	.069
	WSS Magical Ideation Total	4.2 (2.7)	4.5 (3.2)	-.385	.350
	WSS Perceptual Aberration Total	2.5 (2.9)	2.9 (3.3)	-.605	.258
	CAARMS Unusual Thinking Content	5.3 (1.2)	4.9 (1.9)	1.073	.143
	CAARMS Disorganized Speech	2.2 (1.7)	1.6 (1.9)	1.392	.083
	CAARMS Non Bizarre Ideas	4.3 (2.4)	4.9 (1.9)	-1.187	.119
	CAARMS Perceptual Abnormalities	3.6 (2.4)	3.2 (2.5)	.770	.215

2.3 GMV differences between groups

We also wanted to understand the neurobiology that was driving the classifications and misclassifications in our models. We therefore performed whole brain GMV comparisons a) between the whole ROP group and the whole ROD group b) between subjects from all groups who were classified as ROP (regardless of whether it was a correct classification or a misclassification) and c) between Pure subjects and co-morbid subjects. In figure S5 (a,b,c) we can see the GMV differences between the whole of the ROP group and the whole of the ROD group. The ROP group had increased GMV in the superior and inferior frontal gyrus whereas decreased GMV areas were detected in the left cerebellum, the Supramarginal gyrus, and the inferior temporal gyrus when compared to the ROP group (Table S11). Furthermore we compared subjects who were classified as ROP to subject who were classified as ROD according to our imaging model (figure S6 (a,b,c)). Subjects who were classified as ROP showed increased GMV in the insula, the caudate, and the precuneus among others, whereas they showed decreased GMV in the lingual gyrus and the thalamus (Table S12). Finally we performed a whole brain GMV analysis between patients who were in Pure groups (both ROP and ROD) and patients who were in co-morbid groups (both ROP+D and ROD+P) (figure s7 (a,b,c)). Pure subjects showed

pronounced GMV increases in the left cerebellum and the middle occipital gyrus and decreased GMV in the superior frontal gyrus and the lingual gyrus (Table S13).

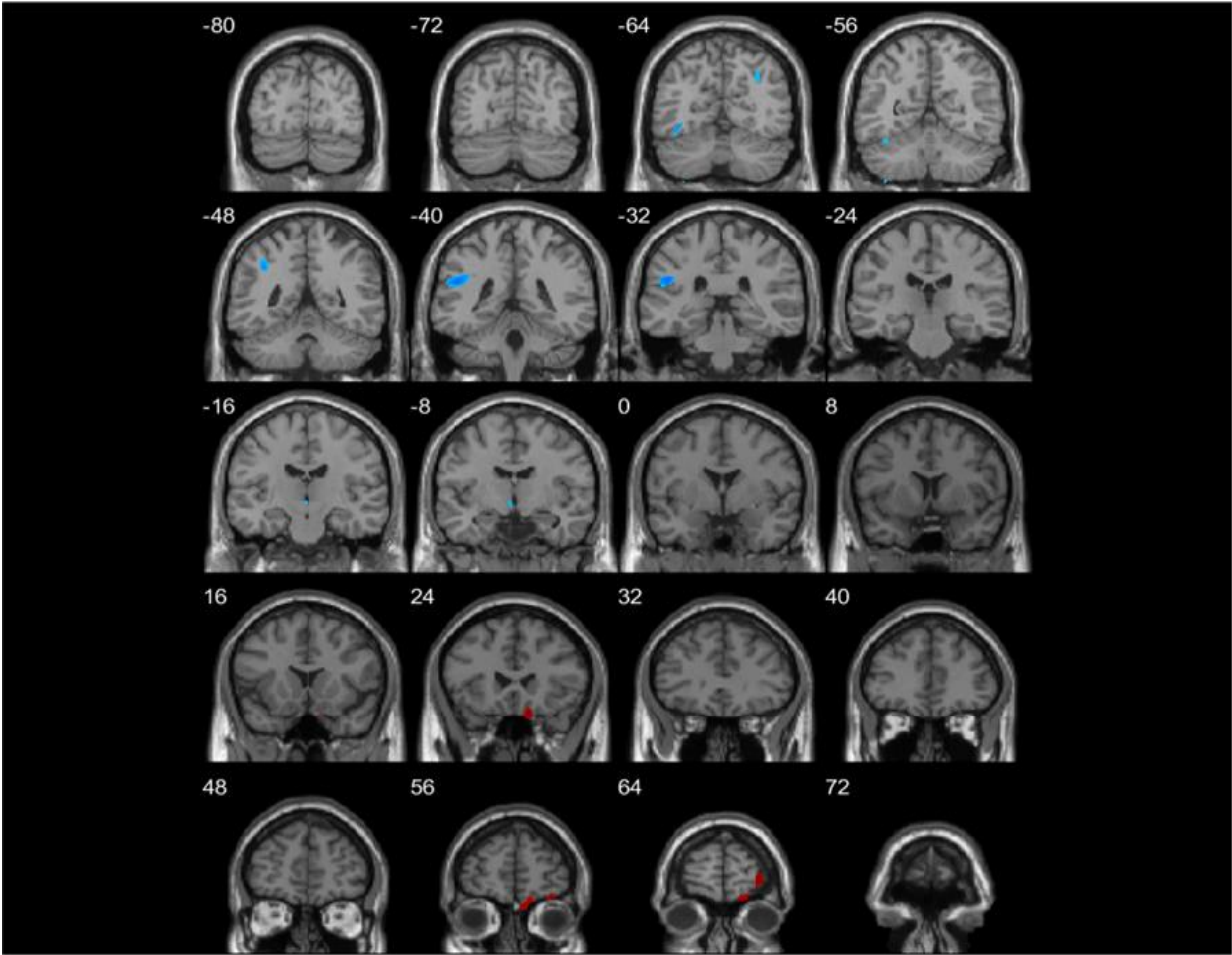


Figure S5a. Whole ROP Group vs Whole ROD Brain Differences

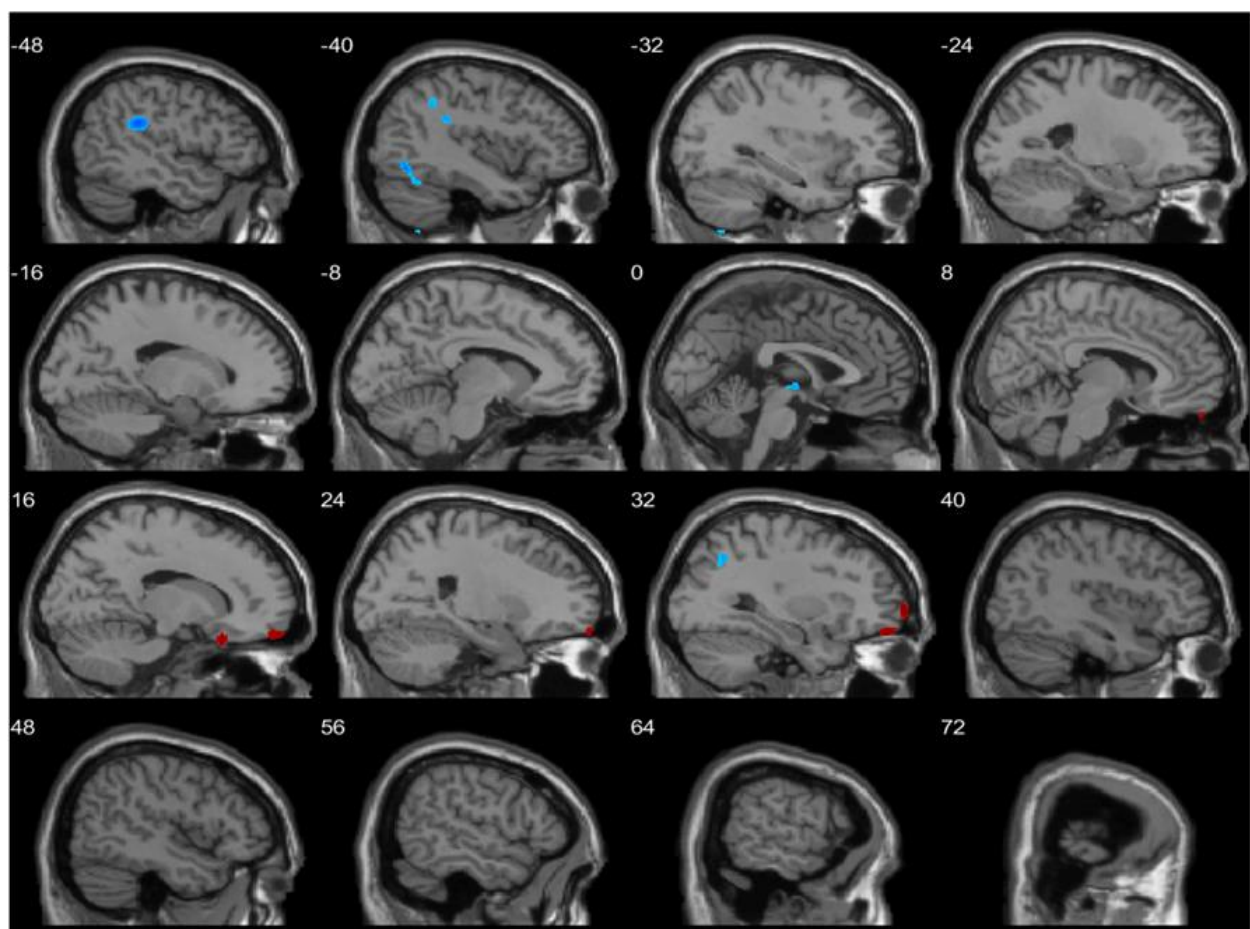


Figure S5b. Whole ROP Group vs Whole ROD Brain Differences

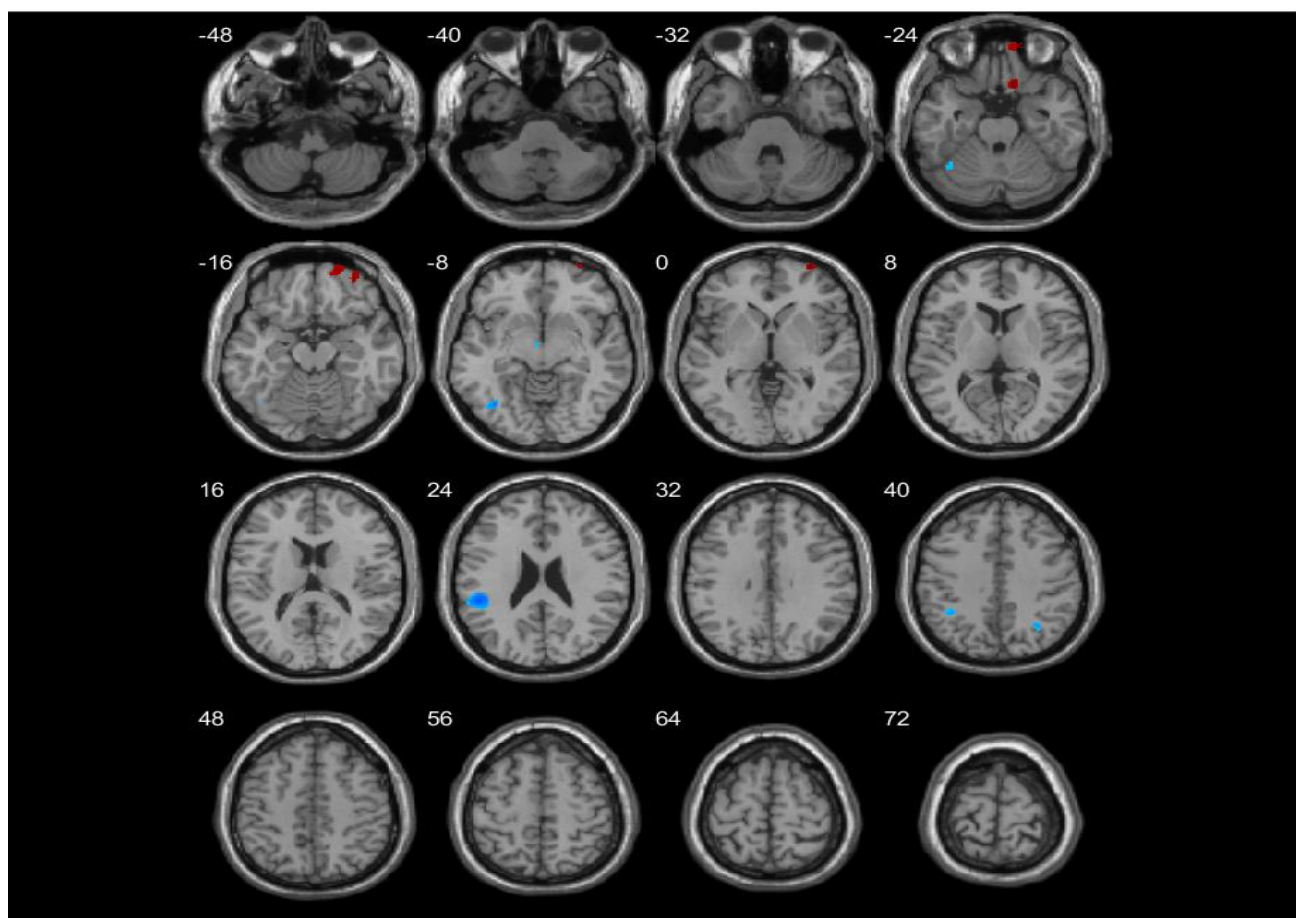


Figure S5c. Whole ROP Group vs Whole ROD Brain Differences

Table S11. Significant clusters from the Whole ROP Group vs Whole ROD Brain Differences analysis. Red=ROP greater GMV than ROD, Blue=ROP less GMV than ROD

Clusters ROP vs ROD

<u>Brain Area</u>	<u>Cluster Peak Voxel MNI Coordinates</u>	<u>P Value, Peak Intensity, Cluster Size</u>
Superior Frontal Gyrus	20 61 -19	p(FDR)=0.013, T=3.64, k=300
Inferior Frontal Gyrus	14 22 -25	p(FDR)=0.057, T=3.84, k=154
Left Cerebellum	-38 -66 -11	p(FDR)=0.044, T=-4.16, k=208
Supramarginal Gyrus	-50 -36 24	p(FDR)=0.001, T=-4.17, k=548
Inferior Temporal Gyrus	-22 3 -55	p(FDR)=0.001, T=-4.55, k=682

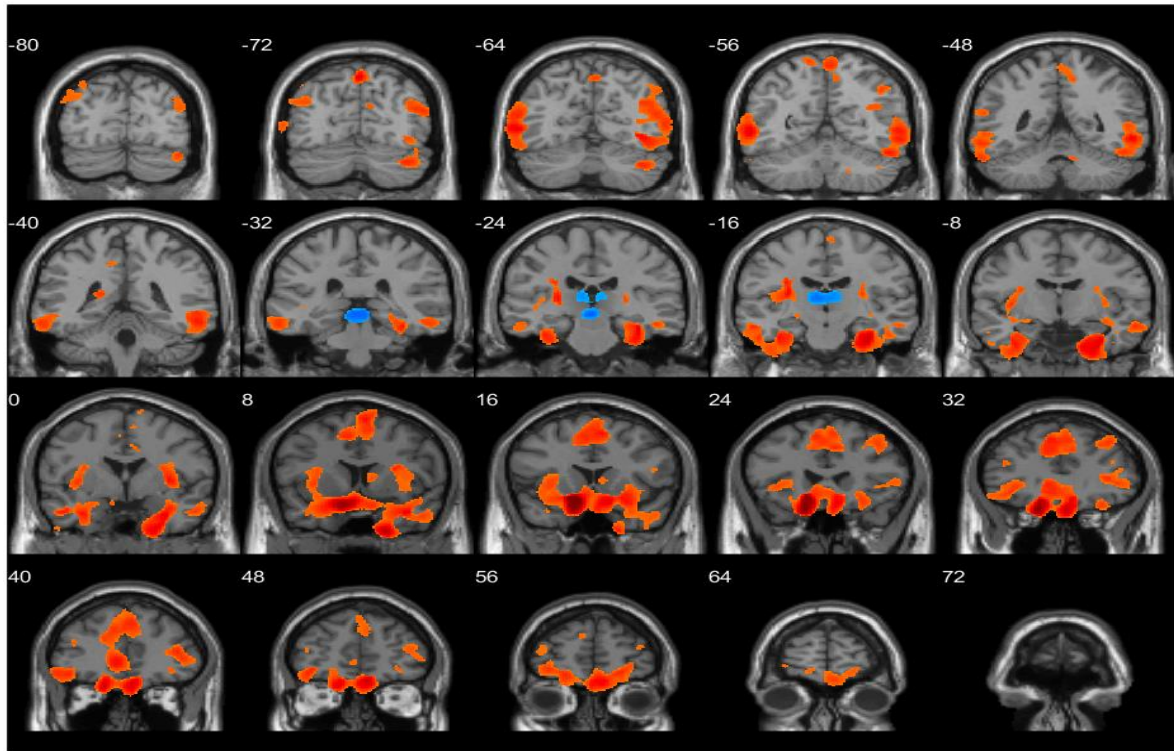


Figure S6a. Subjects Classified as ROP vs Subjects who were classified as ROD

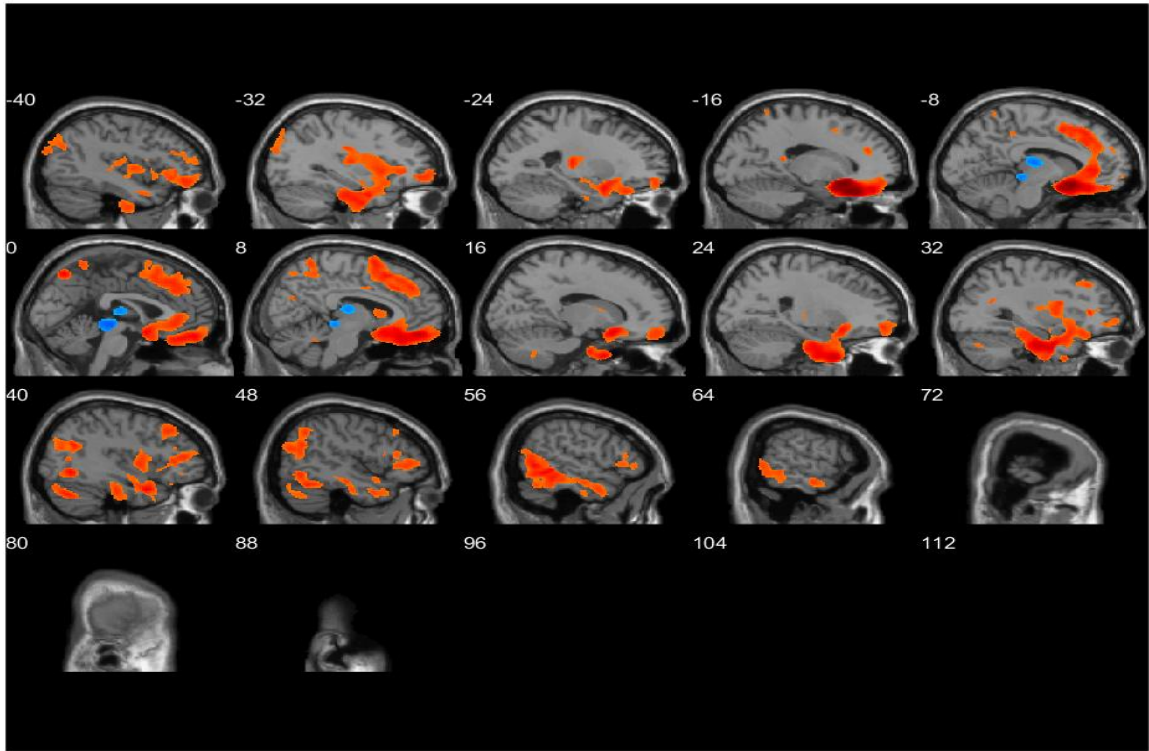


Figure S6b. Subjects Classified as ROP vs Subjects who were classified as ROD

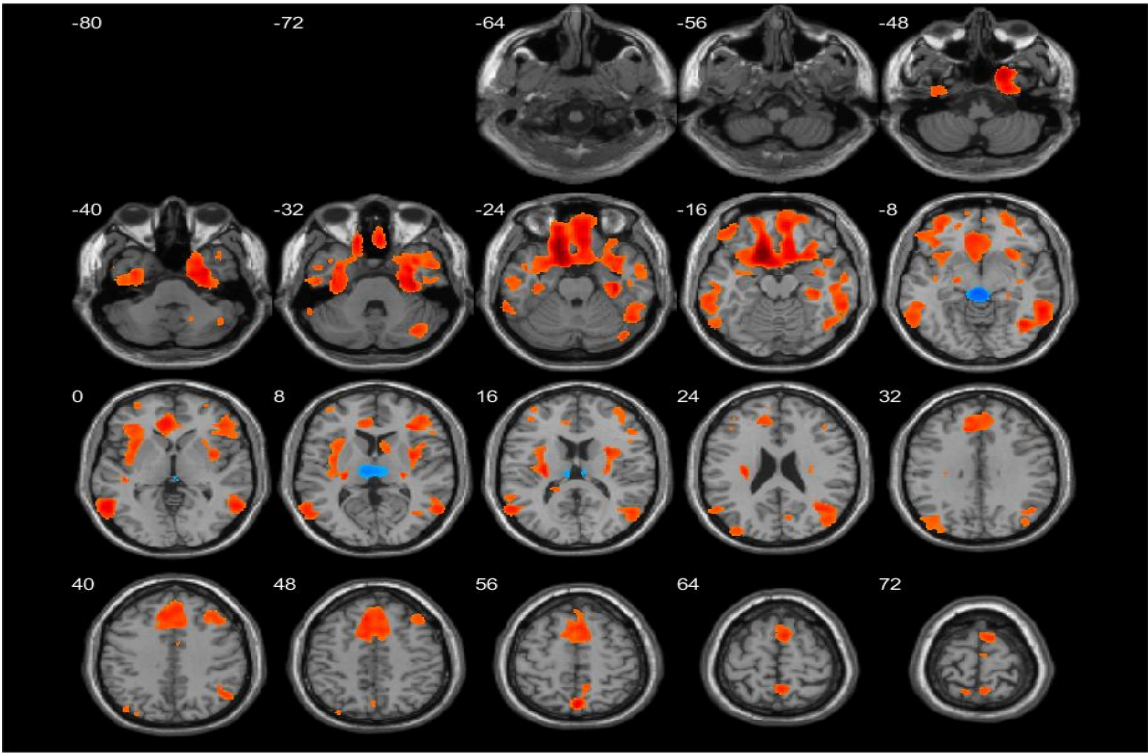


Figure S6c. Subjects Classified as ROP vs Subjects who were classified as ROD

Table S12. Subjects from all Groups Classified as ROP vs Subjects from all Groups Classified as ROD According to the Imaging Model. Red=ROP greater GMV than ROD, Blue=ROP less GMV than ROD

Clusters Classified as ROP vs Classified as ROD

<u>Brain Area</u>	<u>Cluster Peak Voxel MNI Coordinates</u>	<u>P Value, Peak Intensity, Cluster Size</u>
Middle Temporal Gyrus/Inferior Frontal Gyrus/Insula	-16 18 -13	p(FDR)<0.001, T=6.53, k=61340
Caudate	11 9 6	p(FDR)=0.057, T=3.58, k=288
Middle Frontal Gyrus	-40 34 21	p(FDR)=0.057, T=3.63, k=273
Superior Occipital Gyrus/Precuneus	-35 -86 24	p(FDR)<0.001, T=3.78, k=1117
Middle Frontal Gyrus	40 28 43	p(FDR)<0.001, T=3.98, k=826
Precuneus	1 -72 55	p(FDR)<0.001, T=5.12, k=1133

Lingual Gyrus	-2 -27 -8	p(FDR)<0.001, T=-4.35, k=888
Thalamus	4 -18 10	p(FDR)<0.001, T=-4.06, k=992

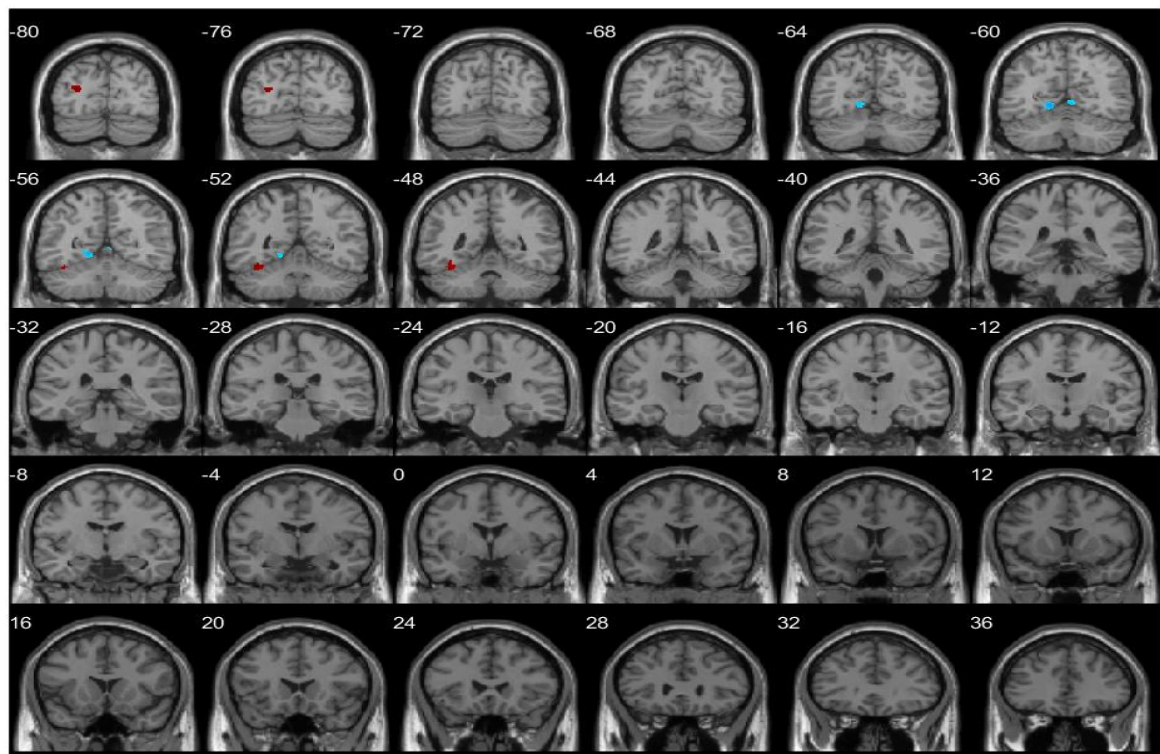


Figure S7a. Subjects in Pure Groups vs Subjects in Co-Morbid Groups

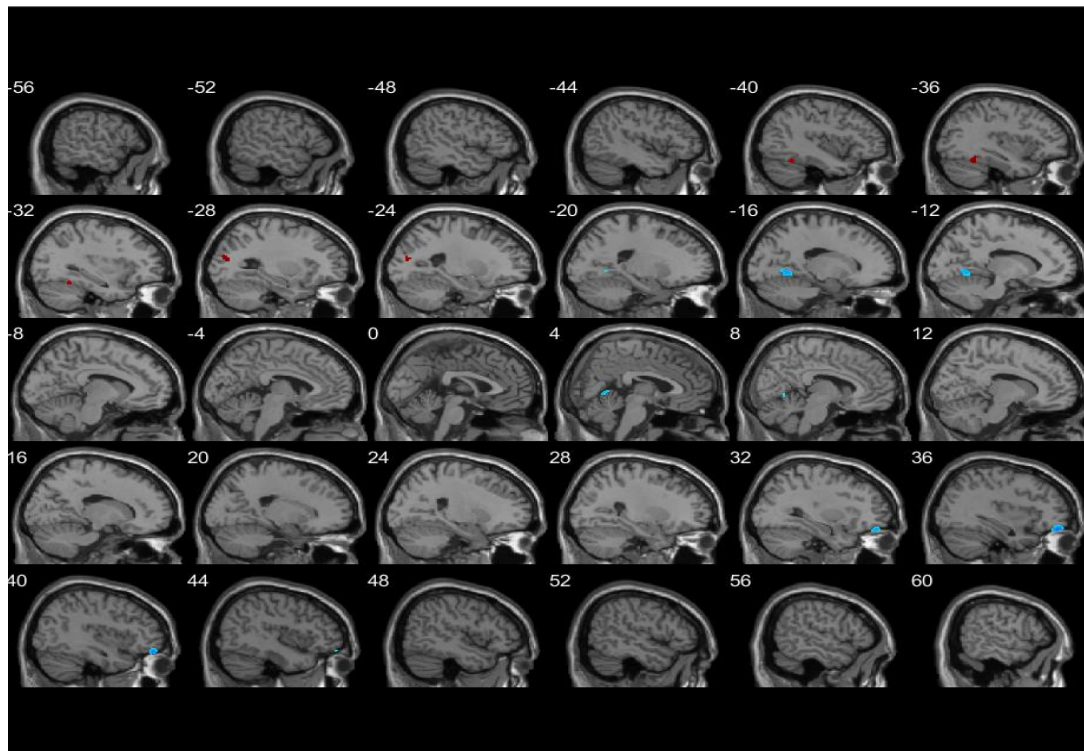


Figure S7b. Subjects in Pure Groups vs Subjects in Co-Morbid Groups

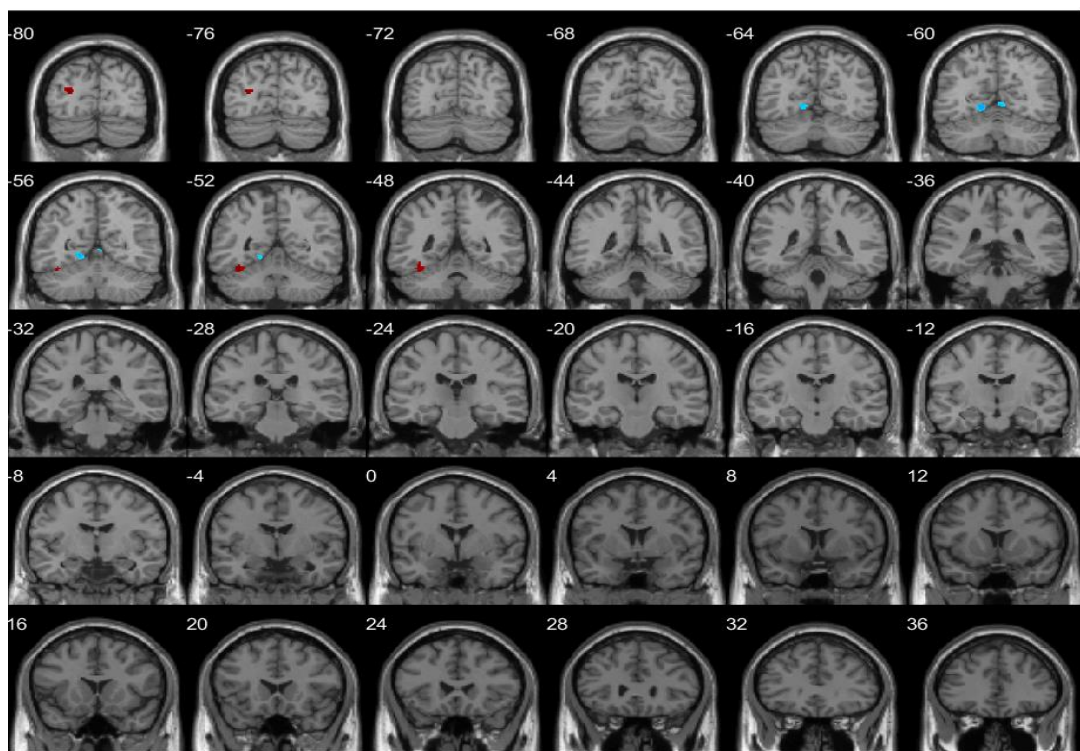


Figure S7c. Subjects in Pure Groups vs Subjects in Co-Morbid Groups

Table S13. Subjects in Pure Groups (Both ROP and ROD) vs Subjects in Co-Morbid Groups (Both ROP Depressed and ROD Psychotic) Brain Differences. Red=ROP greater GMV than ROD, Blue=ROP less GMV than ROD

<i>Clusters Pure vs Co-Morbid</i>		
<u>Brain Area</u>	<u>Cluster Peak Voxel MNI Coordinates</u>	<u>P Value, Peak Intensity, Cluster Size</u>
Left Cerebellum	-38 -51 -21	p(FDR)=0.080, T=3.80, k=155
Middle Occipital Gyrus	-27 -80 17	p(FDR)=0.080, T=3.70, k=120
Superior Frontal Gyrus	37 51 -20	p(FDR)=0.013, T=-4.59, k=304
Lingual Gyrus	-16 -57 -4	p(FDR)=0.018, T=-3.85, k=230

2.5. Decision Scores Group Differences

We wanted to test the differences of the decision scores between the groups (ROP and ROD as well as ROP+D and ROD+P). In the pure clinical model the mean decision score for the ROP group was .4108 ($SD=0.805$) and for the ROD group -.7760 ($SD=0.726$). An independent samples t-test showed that the difference between the two groups was significant ($t=8.177$, $df=126$, $p<.0001$). In the pure neuroanatomical model the mean decision score for the ROP group was -.0579 ($SD=0.227$) and for the ROD group -.2096 ($SD=0.178$). An independent samples t-test showed that the difference between the two groups was significant ($t=4.039$, $df=126$, $p=.0001$). In the pure combined model the mean decision score for the ROP group was .6988 ($SD=1.121$) and for the ROD group -1.019 ($SD=1.012$). An independent samples t-test showed that the difference between the two groups was significant ($t=8.490$, $df=126$, $p<.0001$).

In the applied clinical model the mean decision score for the ROP group was -.2507 ($SD=0.698$) and for the ROD group -.6742 ($SD=0.701$). An independent samples t-test showed that the difference between the two groups was significant ($t=2.686$, $df=115$, $p=.005$). In the applied neuroanatomical model the mean decision score for the ROP group was -.1619

($SD=1.859$) and for the ROD group $-.1217$ ($SD=0.201$). An independent samples t-test showed that the difference between the two groups was not significant ($t=-.942$, $df=115$, $p=.187$). Finally, in the applied combined model the mean decision score for the ROP group was $-.2798$ ($SD=1.003$) and for the ROD group $-.8769$ ($SD=1.008$). An independent samples t-test showed that the difference between the two groups was significant ($t=2.635$, $df=115$, $p=.006$).

2.6. Decision Scores Regression with 9 Month Functional Outcome

Finally we wanted to test whether decision scores were related to 9 month global functioning outcome using the Global Functioning: Social (GFS) and Global Functioning Role (GFR) scales. Good functioning was interpreted as a score on either scale of more than 7, and the impaired functioning was interpreted as score on either scale of 7 or less (Koutsouleris et al., 2018b). A logistic regression was performed to determine the likelihood that patients had an impaired functional outcome at 9 months using decision scores from our SVM models as predictors. None of the models was statistically significant. GFS model using pure SVM model decision scores: $\chi^2(3) = 2.259$, $p = .520$; GFR model using pure SVM model decision scores: $\chi^2(3) = 3.638$, $p = .303$; GFS model using comorbid SVM model decision scores: $\chi^2(3) = 7.179$, $p = .066$; GFR model using comorbid SVM model decision scores: $\chi^2(3) = 3.353$, $p = .340$. Table S12 contains information from the logistic regression models.

Table S14. Odds Ratios (OR) for Impaired Functional Outcomes for Decision Scores from the SVM models

			Beta/OR (95% Confidence Intervals)					
Global Functioning Scale	Outcome	No.	Pure Clinical & Neurocognitive Model	Pure GMV Model	Pure Stacking Model	Comorbid Clinical & Neurocognitive Model	Comorbid GMV Model	Comorbid Stacking Model
GFR	Impaired	Pure: 70 Comorbid: 84						
	Good	Pure: 58 Comorbid: 33	-1.372/.254 (.008-7.812)	-.956/.384 (.063-2.348)	.814/.2.256 (.200-25.406)	23.189/1178.0 (.003-4263.0)	.319/1.375 (.155-12.175)	-15.961/.000 (.000-61.799)
GFS	Impaired	Pure: 69 Comorbid: 77						
	Good	Pure: 59 Comorbid: 40	1.474/4.365 (.138-137.5)	-1.023/.359 (.060-2.144)	-.912/.402 (.035-4.582)	34.917/1460.0 (251.3-8478.0)	.564/1.758 (.215-14.389)	-24.104/.000 (.000-.025)

Acknowledgements:

Author Contributions: Mr. Lalousis, Dr. Koutsouleris, and Prof. Upthegrove had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors reviewed, revised, and approved the final version of the manuscript.

Acquisition and analysis of data: Lalousis, Wood, Chisholm, Griffiths, Borgwardt, Brambilla, Pantelis, Bonivento, Dwyer, Ferro, Haidl, Rosen, Schmidt, Koutsouleris, and Upthegrove

Drafting of the manuscript: Lalousis, Wood, Schmaal, Reniers, Koutsouleris, and Upthegrove

Critical revision of the manuscript for important intellectual content: Lalousis, Wood, Schmaal, Bertolino, Borgwardt, Brambilla, Kambeitz, Lencer, Pantelis, Ruhrmann, Salokangas, Schultze-Lutter, Dwyer, Schmidt, Meisenzahl, Koutsouleris, and Upthegrove

Statistical analysis and interpretation of data: Lalousis, Wood, Schmaal, Koutsouleris and Upthegrove

Obtained funding and designed the study: Wood, Bertolino, Borgwardt, Brambilla, Kambeitz, Lencer, Pantelis, Ruhrmann, Salokangas, Meisenzahl, Koutsouleris, and Upthegrove

Administrative, technical, or material support: Chisholm, Borgwardt, Brambilla, Haidl, Rosen, Schmidt, Meisenzahl, Koutsouleris, and Upthegrove

Supervision: Wood, Schmaal, Reniers, Borgwardt, Brambilla, Schultze-Lutter, Koutsouleris, and Upthegrove

#The PRONIA consortium:

The authors listed here performed the screening, recruitment, rating, examination, and follow-up of the study participants. They were involved in implementing the examination protocols of the study, setting up its IT infrastructure, and organizing the flow and quality control of the data analyzed in this manuscript between the local study sites and the central study database.

Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany

Linda Betz, Anne Erkens, Eva Gussmann, Shalaila Haas, Alkomiet Hasan, Claudius Hoff, Ifrah Khanyaree, Aylin Melo, Susanna Muckenhuber-Sternbauer, Janis Köhler, Ömer Öztürk, Nora Penzel, David Popovic, Adrian Rangnick, Sebastian von Saldern, Rachele Sanfelici, Moritz Spangemacher, Ana Tupac, Maria Fernanda Urquijo, Johanna Weiske, Antonia Wosgien

Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany

Karsten Blume, Dominika Gebhardt, Nathalie Kaiser, Ruth Milz, Alexandra Nikolaides, Mauro Seves, Silke Vent, Martina Wassen

Department of Psychiatry (Psychiatric University Hospital, UPK), University of Basel, Switzerland

Christina Andreou, Laura Egloff, Fabienne Harrisberger, Claudia Lenz, Letizia Leanza, Amatya Mackintosh, Renata Smieskova, Erich Studerus, Anna Walter, Sonja Widmayer

Institute for Mental Health & School of Psychology, University of Birmingham, United Kingdom

Chris Day, Mariam Iqbal, Mirabel Pelton, Pavan Mallikarjun, Alexandra Stainton, Ashleigh Lin

Department of Psychiatry, University of Turku, Finland

Alexander Denissoff, Anu Ellilä, Tiina From, Markus Heinimaa, Tuula Ilonen, Päivi Jalo, Heikki Laurikainen, Antti Luutonen, Akseli Mäkela, Janina Paju, Henri Pesonen, Reetta-Liina Säilä, Anna Toivonen, Otto Turtonen

General Electric Global Research Inc., USA

Ana Beatriz Solana, Manuela Abraham, Nicolas Hehn, Timo Schirmer

Workgroup of Paolo Brambilla, University of Milan, Italy:

**Department of Neuroscience and Mental Health, Fondazione IRCCS Ca' Granda
Ospedale Maggiore Policlinico, University of Milan, Milan, Italy:** Carlo Altamura, Marika
Belleri, Francesca Bottinelli, Adele Ferro, Marta Re

Programma2000, Niguarda Hospital, Milan: Emiliano Monzani, Maurizio Sberna

San Paolo Hospital, Milan: Armando D'Agostino, Lorenzo Del Fabro

Villa San Benedetto Menni, Albese con Cassano (CO): Giampaolo Perna, Maria Nobile,
Alessandra Alciati

Workgroup of Paolo Brambilla at the University of Udine, Italy

Department of Medical Area, University of Udine, Udine, Italy: Matteo Balestrieri,
Carolina Bonivento, Giuseppe Cabras, Franco Fabbro

IRCCS Scientific Institute "E. Medea", Polo FVG, Udine: Marco Garzitto, Sara Piccin

Appendix C

Neurobiologically Based Stratification of Recent Onset Depression and Recent Onset Psychosis: Identification of Two Distinct Transdiagnostic Phenotypes

Paris Alexandros Lalousis, MSc^{1,2}; Lianne Schmaal, PhD^{3,4}; Stephen J. Wood, PhD^{1,3,4}; Renate L.E.P Reniers, PhD^{1,2,5}; Nicholas M. Barnes, PhD⁵; Katharine Chisholm, PhD^{1,6}; Sian Lowri Griffiths, PhD^{1,2}; Alexandra Stainton, PhD^{3,4}; Junhao Wen, PhD⁷; Gyujoon Hwang, PhD⁷; Christos Davatzikos, PhD⁷; Christina Andreou, MD⁸; Carolina Bonivento, PhD^{9,10}; Udo Dannlowski, MD¹¹; Adele Ferro, PhD^{9,10}; Theresa Liechtenstein, MD¹²; Anita Riecher-Rössler, MD⁸; Georg Romer, MD⁹; Marlene Rosen, PhD¹²; Alessandro Bertolino, MD¹³; Stefan Borgwardt, MD^{8,14}; Paolo Brambilla, MD^{9,10}; Joseph Kambeitz, MD¹⁵; Rebekka Lencer, MD, PhD^{9,12}; Christos Pantelis, MB BS, MD, MRCPsych, FRANZCP¹⁶; Stephan Ruhrmann, MD⁹; Raimo K. R. Salokangas, MD, PhD, MSc¹⁷; Frauke Schultze-Lutter, PhD^{18,19,20}; André Schmidt, PhD⁹; Eva Meisenzahl, MD¹⁸; Nikolaos Koutsouleris, MD¹⁵; Dominic Dwyer, PhD^{15*}; Rachel Upthegrove, MBBS FRCPsych, PhD^{1,2,21*}, for the PRONIA Consortium#

* Joint Senior Authors

#See PRONIA consortium author list

Author Affiliations:

- 1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom
- 2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom
- 3 -Orygen, Parkville, Australia
- 4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia
- 5 -Institute for Clinical Sciences, University of Birmingham, United Kingdom

- 6 -Department of Psychology, Aston University, United Kingdom
- 7 -Perelman School of Medicine University of Pennsylvania, United States of America
- 8 -Department of Psychiatry (UPK), University of Basel, Basel, Switzerland
- 9 -Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy
- 10 - Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy
- 11 –Institute for Translational Psychiatry, University of Münster, Münster, Germany
- 12 -Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany
- 13 -Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy
- 14 -Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany
- 15 -Department of Psychiatry and Psychotherapy, Ludwig Maxmilians University, Munich, Germany
- 16 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia
- 17 -Department of Psychiatry, University of Turku, Turku, Finland
- 18 -Department of Psychiatry and Psychotherapy, University of Düsseldorf, Düsseldorf, Germany
- 19 -Department of Psychology and Mental Health, Faculty of Psychology, Airlangga University, Surabaya, Indonesia
- 20 - University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland
- 21 -Birmingham Early Interventions Service, Birmingham Women's and Children's NHS Foundation Trust

Corresponding Author: Paris Alexandros Lalousis, BSc, MSc, School of Psychology, University of Birmingham, 52 Pritchatts Road, Birmingham, B15 2SA, United Kingdom



2. Supplementary Methods

PRONIA recruitment infrastructure

The 981 study participants (155 individuals with ROP, 147 individuals with ROD, and 275 HC from the discovery sample as well as 129 individuals with ROP, 123 individuals with ROD, and 152 HC from the replication sample) analyzed in the present study were recruited following a standardized recruitment and ascertainment protocol (see **Figure S1** and **Table S3**). The observational study protocol involved follow-up examinations every three months after the index ascertainment and was implemented by the following 7 PRONIA sites:

Table S3: Characteristics of the recruiting institutions in the PRONIA consortium (Directly reproduced from a previous study(Koutsouleris et al., 2018b)).

PRONIA Site	Institution Name	Country	Type of Service	Catchment Population	Screening population / year
Munich	Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich	DE	Academic outpatient services including specialized service for early recognition of psychosis; tertiary care academic hospital	1,200,000	700
Basel	Department of Psychiatry and Psychotherapy, University of Basel	CH	Academic inpatient and outpatient services including specialized service for early recognition and intervention of psychosis; tertiary care academic hospital	500,000	200
Milan Niguarda	Department of Pathophysiology and Transplantation, University of Milan. Four recruitment hospitals: Niguarda, Policlinico, San Paolo, Villa San Benedetto Menni in Albese con Cassano	IT	Psychiatric outpatient services including specialized services for early recognition of psychosis and persons at high risk; Academic hospital, providing psychiatric inpatient services, psychiatric outpatient services and local services;	600,000	1,000
Cologne	Department of Psychiatry and Psychotherapy, University of Cologne	DE	Academic outpatient services including specialized service for early recognition of psychosis; tertiary care academic hospital	1,000,000	600
Birmingham	The University of Birmingham	UK	Academic specialised Early Intervention Service for Psychosis covering Birmingham and Solihull. Community and Inpatient	1,200,000	800
Turku	Department of Psychiatry, University of Turku	FI	Psychiatric outpatient and hospital services responsible for treatment of psychiatric patients in their catchment areas in the South-Western Finland.	284,000	2,300

Udine	Department of Psychiatry, University of Udine	IT	Psychiatric outpatient services, academic hospital and local services. Tertiary care neuropsychiatric service	600,000	500
-------	--	----	---	---------	-----

Upon study enrolment, the participants were pseudonymized twice, locally at each site and centrally at the level of the PRONIA portal. The PRONIA portal consists of a multi-user database hosting the clinical and neurocognitive information, and defaced MR images obtained from the study participants. The data are organized into digital questionnaires, visits, and cases. The portal provides the case managers with a controlled web-based interface to enter and upload the different data into the respective questionnaires. Furthermore, the PRONIA consortium has implemented a PRONIA@home mobile device interface that allows the study participants to securely log into the portal and fill out the self-rating questionnaires of given visit. Upon completion of the data entry across all questionnaires of given visit, the data are checked by an automatic quality control procedure which executes approximately 1600 data integrity and dependency rules. These rules include (1) basic checking of missing data and data ranges, (2) checking of dependency within one questionnaire, (3) dependencies between two questionnaires within one visit, and (4) dependencies between two consecutive visits (such as consistency of dates). Detected errors are fed back to the respective case managers allowing for a manual correction of the respective issues. This process is re-iterated until the quality of the clinical questionnaires in the given visit is sufficient for the entire visit to be locked.

PRONIA study design and examination instruments

A comprehensive battery of ascertainment tools was used within a longitudinal observational study design to generate a multi-modal phenotypic profile of each study participant (see **Figure S1** and **Table S2**). The clinical part of the battery compiled questionnaires that capture sociodemographic, somatic, environmental, diagnostic, psychopathological, functional and quality-of-life related variables in the PRONIA study population. This battery was complemented by multi-domain neurocognitive and neuroimaging examinations as well as blood sampling for later genetic characterization, which were carried out at the baseline and 9-month follow-up timepoints.

Table S4: Clinical and neurocognitive examinations performed in the CHR, ROD, ROP, and HC groups during the 18-month follow-up period of the study (Directly reproduced from a previous study(Koutsouleris et al., 2018b)). Clinical assessment types: OR Observer-based rating instrument, **SR** Self-rating-based instrument. **Examination timepoints: T0** Baseline examination,

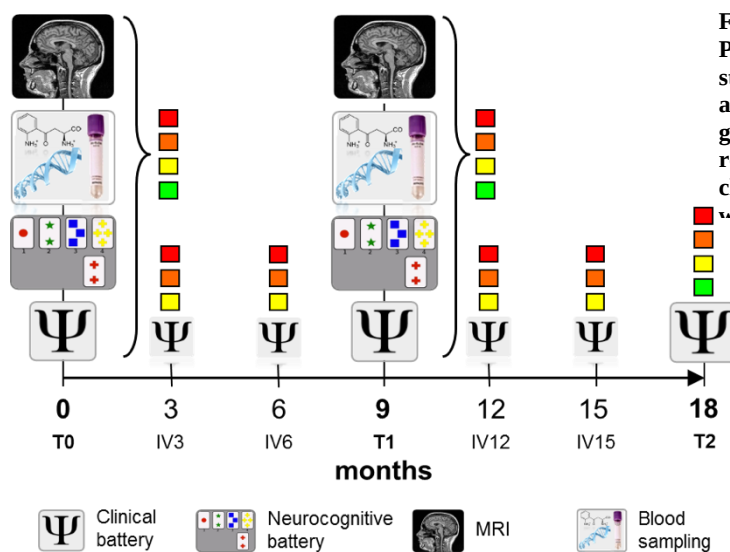


Figure S2: Observational study design of PRONIA (Directly reproduced from a previous study¹). Colored boxed indicate type of assessment / visits conducted in each of the study groups: Healthy controls (green), patients with recent-onset depression (yellow), persons with a clinical high-risk for psychosis (orange), patients with recent-onset psychosis (red).

IV3/IV6/IV12/IV15 3, 6, 12, 15-month examinations conducted only in the clinical study participants, T1 9-month examination, T2 18-month follow-up examination.

Observer-based instruments: CAARMS Comprehensive Assessment of the At-Risk Mental States(Yung et al., 2005), *CHR Criteria* Clinical High-Risk criteria summary questionnaire, *FROGS* Functional Remission in General

Schizophrenia(Llorca et al., 2009), *GAF* Global Assessment of Functioning, *GF:S/R* Global Functioning: Social / Role(Cornblatt et al., 2007), *PANSS* Positive and Negative Symptom Scale(Kay et al., 1987), *PAS* Premorbid Adjustment Scale(Alvarez et al., 1987), *SANS* Scale for the Assessment of Negative Symptoms(Andreasen, 1989), *SCID-IV Screening/Summary* Structured Clinical Interview for DSM-IV(First, Michael B et al., 2002), *SIPS* Standardized Interview for the Assessment of Prodromal Symptoms (modified version 5.0)(McGlashan et al., 2001), *SPI-A [COGDIS/COPER]* Schizophrenia Proneness Instrument [Cognitive disturbances (COGDIS) / Cognitive-Perceptual (COPER) disturbances](F. Schultze-Lutter et al., 2007), Transition Criteria Interval questionnaire for the assessment of transition criteria, *UHR - Schizotypy*, *Genetic Risk* Interview for the Assessment of Schizotypal personality traits, and familial risk for psychosis. **Self-rating instruments:** *BDI-II* Beck Depression Inventory II(A. Beck et al., 1996), *CISS-24* Coping Inventory for Stressful Situations – 24 items(Endler, 2004), *CTQ* Childhood Trauma Questionnaire(Bernstein et al., 1994), *EHF-SR* Edinburgh Handedness Inventory – Short Version(Veale, 2014), *EDS* Everyday Discrimination Scale – Modified Version(D. R. Williams, n.d.), *LEE* Level of Expressed Emotions(Cole & Kazarian, 1988), *MSPSS* the Multidimensional Scale for Perceived Social Support(Zimet et al., 1988), *NEO-FFI* NEO Five Factor Inventory of Personality Traits(Costa et al., 1992), *RSA* Resilience Scale for Adults(Friborg et al., 2003), *SPIN* Social Phobia

Inventory(Connor et al., 2000), *WHO-QOL-BREF* WHO Quality of Life Questionnaire-Brief Version(Health, 1996). Neurocognitive tests: *CPT-IP* Continuous-Performance Test-Identical Pairs (adapted tablet version)(Cornblatt et al., 1988), *DANVA* Diagnostic Analysis of Non-Verbal Accuracy 2 (adapted tablet version)(Nowicki & Duke, 2001), *DS* Auditory Digit Span (Forward/Backward) adapted from the PEBL battery, *DSST* Digit-Symbol-Substitution Test from the BACS battery, *ROCF* Rey-Osterrieth complex figure(Rey, 1941), *SAT* Salience Attribution Task (adapted version)(Roiser et al., 2010), *SOPT* self-ordered pointing task (adapted version)(Petrides & Milner, 1982), *TMT-A/B* Trail-Making Test A and B(Reitan, 1958), *VF* phonemic/semantic verbal fluency test.

Instrument	Form	Screening		T0		IV3	IV6	T1		IV12	IV15	T2	
		PAT	HC	PAT	HC	PAT	PAT	PAT	HC	PAT	PAT	PAT	HC
General Data	OR	X	X					X	X			X	X
Reasons for Referral	OR	X											
Treatment Documentation	OR	X	X			X	X	X	X	X	X	X	X
Somatic state and Health History	OR	X	X					X	X			X	X
SPI-A COGDIS/COPER	OR	X	X			X	X	X	X	X	X	X	X
SIPS positive symptoms	OR	X	X			X		X	X			X	X
CAARMS	OR	X	X			X		X	X			X	X
GAF	OR	X	X			X		X	X			X	X
UHR – Schizotypy, Genetic Risk	OR	X	X			X		X	X			X	X
CHR Criteria	OR	X	X					X	X			X	X
Transition Criteria	OR					X	X			X	X		
SCID-IV Screening	OR	X	X					X	X			X	X
SCID-IV Summary	OR	X	X					X	X			X	X
Demographic and Biographic Data	OR			X	X			X	X			X	X
PAS	OR			X	X			X				X	
SPI-A	OR			X	X			X				X	
SIPS negative, disorganized and general symptoms	OR			X	X			X				X	
PANSS	OR			X		X	X	X		X	X	X	
SANS	OR			X				X				X	
Chart of Life Events	OR			X	X	X	X	X	X	X	X	X	X
FROGS	OR			X				X				X	
GF: Social & Role	OR			X	X	X	X	X	X	X	X	X	X
Prognostic evaluation	OR			X				X				X	
MSPSS	SR			X	X			X	X			X	X
RSA	SR			X	X			X	X			X	X
CISS 24	SR			X	X			X	X			X	X
SPIN	SR			X	X			X	X			X	X
BDI-II	SR			X	X	X	X	X	X	X	X	X	X
WHO-QOL-BREF	SR			X	X			X	X			X	X
EH-SR	SR			X	X								
LEE	SR			X	X			X	X			X	X
Wisconsin Scales	SR			X	X								
EDS	SR			X	X								
Bullying Scale T0	SR			X	X								
CTQ	SR			X	X								
NEO-FFI	SR			X	X								
DS backward (BACS)	NPT			X	X			X	X				
DS forward (BACS)	NPT			X	X			X	X				
CPT-IP (BACS)	NPT			X	X			X	X				
DANVA	NPT			X	X			X	X				
DSST	NPT			X	X			X	X				
RAVLT	NPT			X	X			X	X				
ROCF	NPT			X	X			X	X				
SAT	NPT			X	X			X	X				
SOPT	NPT			X	X			X	X				
TMT-A	NPT			X	X			X	X				
TMT-B	NPT			X	X			X	X				
VF phonetic	NPT			X	X			X	X				
VF semantic	NPT			X	X			X	X				
WAIS-III	NPT			X	X			X	X				

MRI harmonization and data acquisition

When setting up the PRONIA study, we decided to generate a MRI database that would represent the MR scanner sequence heterogeneity encountered in clinical real-world. The aim of this strategy was to strengthen the generalizability and clinical applicability of the predictive models developed by our machine learning analyses. Thus, we agreed on a minimal harmonization protocol that required the PRONIA sites to only (1) acquire isotropic or nearly isotropic voxel sizes of preferably 1 mm resolution, (2) set the Field Of View (FOV) parameters accordingly to guarantee the full 3D coverage of the brain including all parts of the cerebellum, and (3) define the relaxation time (TR) and echo time (TE) as well as other imaging parameters in a way that would maximize the contrast between cortical ribbon and the white matter and enhance the signal-to-noise ratio in the images. The sMRI images of six healthy travelling volunteers who were scanned at all sites with the same parameters, were also analysed as part of a calibration study. At every site all the images were visually inspected, automatically defaced and anonymized using an in-house Freesurfer-based script before the data was centralized. **Table S3** lists the parameters defining the structural MR sequences used to examine in the PRONIA discovery sample participants.

Table S3: MR scanner systems and structural MRI sequence parameters used at the respective PRONIA sites (Directly reproduced from a previous study(Koutsouleris et al., 2018b)). .

PRONIA Site	Model	Field Strength	Coil Channels	Flip Angle	TR [ms]	TE [ms]	Voxel Size [mm]	FOV	Slice Number
Munich	Philips Ingenia	3T	32	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Milan Niguarda	Philips Achieva Intera	1.5T	8	12	Shortest (8.1)	Shortest (3.7)	0.93 x 0.93 x 1.0	240 x 240	170
Basel	SIEMENS Verio	3T	12	8	2000	3.4	1.0 x 1.0 x 1.0	256 x 256	176
Cologne	Philips Achieva	3T	8	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Birmingham	Philips Achieva	3T	32	8	8.4	3.8	1.0 x 1.0 x 1.0	288 x 288	175
Turku	Philips Ingenuity	3T	32	7	8.1	3.7	1.0 x 1.0 x 1.0	256 x 256	176
Udine	Philips Achieva	3T	8	12	Shortest (8.1)	Shortest (3.7)	0.93 x 0.93 x 1.0	240 x 240	170

MRI processing pipeline

The manual of the CAT12 toolbox (<http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>) details the processing steps applied to the structural images. These steps consist of:

- (9) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering(Manjón et al., 2008).
- (10) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local variations of intensity distributions as slowly varying spatial functions and thus achieves a homogeneous segmentation across cortical and subcortical structures(Rajapakse et al., 1997).
- (11) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior information of adjacent voxels into the segmentation estimation generated by AMAP(Rajapakse et al., 1997).
- (12) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM) inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP segmentation.
- (13) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the AMAP-generated tissue segments.
- (14) A high-dimensional DARTEL registration of the image to a MNI-template generated from the MRI data of 555 healthy controls in the IXI database (<http://www.braindevelopment.org>).
- (15) The GM maps were then multiplied with the Jacobian determinants that were obtained during registration in order to produce GM volume maps.

(16) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

1.5 ROP and ROD Inclusion Criteria

The general inclusion criteria for the study were: (1) age between 15 and 40 years, (2) sufficient language skills for participation, (3) capacity to provide informed consent/assent. General exclusion criteria were: (1) an IQ below 70, (2) current or past head trauma with loss of consciousness (> 5 minutes), (3) current or past known neurological or somatic disorders potentially affecting structure or functioning of the brain, (4) current or past alcohol dependence, (5) polysubstance dependence within the past six months, and (6) any medical indication against MRI. Recent onset psychosis (ROP) and recent onset depression (ROD) inclusion criteria can be found in the supplement (1.5). ROP participants had to meet the following criteria: 1) DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders, Text Revision) (43) affective or non-affective psychotic episode (lifetime), 2) criteria for DSM-IV-TR affective or nonaffective psychotic episode fulfilled within past 3 months and 3) onset of psychosis within past 24 months. ROD patients had to meet the following criteria: 1) DSM-IV-TR major depressive episode (lifetime), 2) major depressive disorder criteria fulfilled within past three months and 3) duration of first depressive episode no longer than 24 months. Specific ROD exclusion criteria were: (1) a previous episode of DSM-IV-TR major depression prior to the current or recent episode, (2) a duration of the current episode exceeding 24 months, (3) antipsychotic medication exceeding 30 days (cumulative in the past 24 months) with a daily dose rate at or above minimum dosage of the “first episode psychosis” range of DGPPN S3 guidelines (Psychiatrie P und P und N (DGPPN) Deutsche Gesellschaft für, 2006), and (4) any antipsychotic medication within the past three months prior to baseline assessments at or above minimum dosage of the “first episode psychosis”

range of DGPPN S3(Psychiatrie P und P und N (DGPPN) Deutsche Gesellschaft für, 2006)
(See Table S4).

Table S4: DGPPN S3 Guidelines for the treatment of first-episode psychosis and schizophrenia (translated English version of Table 4.1 stated in the short version of the guideline manual available in [https://www.dgppn.de/ Resources/Persistent/a6e04aa47e146de9e159fd2ca1e6987853a055d7/S3_Schizo_Kurzversion.pdf](https://www.dgppn.de/Resources/Persistent/a6e04aa47e146de9e159fd2ca1e6987853a055d7/S3_Schizo_Kurzversion.pdf)) (Directly reproduced from a previous study(Koutsouleris et al., 2018b)). Candidate CHR and ROD patients were excluded if they had received antipsychotic medication (1) for more than 30 cumulative days at or above the minimum target dosage threshold for the treatment of first-episode psychosis, or (2) within the past 3 months before psychopathological baseline assessments at or above the minimum target dosage threshold for the treatment of first-episode psychosis. Abbreviations: DI dosage interval, ²maximum recommended dosage according to prescribing information.

Substance	Recommended starting dosage (mg/d)	DI'	Target dosage first-episode psychosis (mg/d)	Target dosage relapsing schizophrenia (mg/d)	Maximum dosage recommended (mg/d) ²
Atypical Antipsychotics					
Amisulpride	200	(1)-2	100-300	400-800	1200
Aripiprazole	(10)-15	1	15-(30)	15-30	30
Olanzapine	5-10	1	5-15	5-20	20
Quetiapine	50	2	300-600	400-750	750
Risperidone	2	1-2	1-4	3-6-(10)	16
Ziprasidone	40	2	40-80	80-160	160
Typical Antipsychotics					
Fluphenazine	0.4-10	2-3	2.4-10	10-20	20-(40)
Flupentixole	2-10	1-3	2-10	10-60	60
Haloperidole	1-10	(1)-2	1-4	3-15	100
Perazine	50-150	1-2	100-300	200-600	1000
Perphenazine	4-24	1-3	6-36	12-42	56
Pimozide	1-4	2	1-4	2-12	16
Zotepine	25-50	2-(4)	50-150	75-150	450
Zuclopenthixole	2-50	1-3	2-10	25-50	75

1.6 HYDRA

HYDRA is a hybrid between unsupervised clustering and supervised classification methods; it can simultaneously fit maximum margin classification boundaries and elucidate disease subtypes, which is not possible with unsupervised clustering methods or non-linear kernel classifiers. HYDRA investigates different directions of deviation from normal anatomy (as dictated by a healthy control reference sample) with no regard for their magnitude. Patients are grouped into clusters based on the brain regions which are different from the healthy controls and not based on the magnitude by which those regions are different. A number of unique hyperplanes is chosen with the application of Determinantal Point Processes (DPP) which is technique that helps derive the most diverse samples possible. This ensures that identified regions reflect biomarkers that are unique rather than repeated ones with varying magnitudes and helps ensure clusters of patients that are reflective of disease related variability.

1.7 Blood-Based Biomarker Data

Serum: Following a strict protocol, a 9 ml blood sample was collected from consenting participants in S-monovette serum tubes for isolation of serum. Blood samples were drawn in the morning after 12 hours fasting where possible. Samples were inverted 5-6 times and allowed to stand at room temperature for 30 minutes to coagulate. After centrifugation for 10 minutes at 2750g, serum was pipetted in 8 300ul aliquots in cryo vials and stored at -80oC until analysis.

Blood: at the same sampling time, two 9ml fasting whole blood samples were collected in EDTA tubes and stored at -80oC until genomic analysis.

Serum Peripheral Inflammatory Marker Assays:

Samples were transferred to the University of Birmingham Barnes Laboratory in March 2019. There were no freeze-thaw cycles before analysis. Of samples included in inflammatory marker

analysis, 83% were taken before 12 noon and 76% were fasting. 95% of participants were in a sitting position when the blood was drawn. The average length of serum aliquot sample storage until analysis was 48.78 months (median = 45 months). All samples were assayed blind to subject information.

An aliquot from each sample was centrifuged (15 min, 15k x g, 4°C) before analysis using the Luminex platform (Bio-Plex 200 system with Bio-Plex Manager software) using a commercial multiplex kit for IFN γ , IL-1b, IL-1ra, IL-2, IL-4, IL-6, S100B and TNFa (Bio-Techne), and commercial singleplex kits for CRP (Bio-Techne), BDNF (Bio-Techne) and TGFb1 (Bio-Techne) following the manufacturer's instructions.

1.8 Independent Validation

We followed the exact same procedure as described in main manuscript in order to develop a model in the replication sample. In order to mitigate site effects, prior to applying HYDRA, the R version of the ComBat harmonization technique was employed (<https://github.com/Jfortin1/ComBatHarmonization>). To further ensure that disease variance would be retained and that scanner variance would be disentangled from disease variance, ComBat was trained on the replication healthy controls and then ~~applied~~ the estimates that were derived were applied to the replication patients. We used whole brain measures derived from CAT12 from 280 Regions of the neuromorphometrics atlas parcellation from 404 participants with Recent Onset Psychosis (ROP) and Recent Onset Depression (ROD) and Healthy Controls (HC) from the replication sample of the PRONIA study. ROP patients and ROD patients were grouped together into one patient group. HYDRA was trained using a 1000-fold cross-validation method using age, sex, and Total Intracranial Volume (TIV) as covariates, and requesting 2 to 8 clustering solutions. Adjusted Rand Index (ARI) was used to measure cluster stability. The statistical significance of clusters was assessed in two ways.

We assessed whether the obtained subgroups were different to what we would find if there was no disease present by comparing the observed ARI to a null distribution of ARI. We also assessed whether clusters really are present in the dataset or if there is a single Gaussian distribution by employing SigClust (Huang et al., 2015)(<https://github.com/pkimes/sigclust2>). Using the discovery data (280 neuromorphometrics features) we then trained a SVM model which classified patients into either the impaired or preserved cluster in order to develop a tool that allows us to apply our clustering solution to external datasets without the need to rerun HYDRA on the original (discovery) data. We then applied that model to the replication sample data. Details of the machine learning pipeline are described in 1.12.

1.9 External Validation

COBRE Dataset

The Center for Biomedical Research Excellence (COBRE) is contributing raw anatomical and functional MR data from 72 patients with Schizophrenia and 75 healthy controls (ages ranging from 18 to 65 in each group). All subjects were screened and excluded if they had; history of neurological disorder, history of intellectual disability, history of severe head trauma with more than 5 minutes loss of consciousness, history of substance abuse or dependence within the last 12 months. Diagnostic information was collected using the Structured Clinical Interview used for DSM Disorders (SCID).

A multi-echo MPRAGE (MEMPR) sequence was used with the following parameters: TR/TE/TI = 2530/[1.64, 3.5, 5.36, 7.22, 9.08]/900 ms, flip angle = 7°, FOV = 256x256 mm, Slab thickness = 176 mm, Matrix = 256x256x176, Voxel size = 1x1x1 mm, Number of echos = 5, Pixel bandwidth = 650 Hz, Total scan time = 6 min. With 5 echoes, the TR, TI and time to encode partitions for the MEMPR are similar to that of a conventional MPRAGE, resulting in similar GM/WM/CSF contrast.

Rest data was collected with single-shot full k-space echo-planar imaging (EPI) with ramp sampling correction using the intercommissural line (AC-PC) as a reference (TR: 2 s, TE: 29 ms, matrix size: 64x64, 32 slices, voxel size: 3x3x4 mm³).

Slice Acquisition Order:

Rest scan - collected in the Axial plane - series ascending - multi slice mode - interleaved

MPRAGE - collected in the Sag plane - series interleaved - multi slice mode - single shot

The following data are released for every participant:

Resting fMRI

Anatomical MRI

Phenotypic data for every participant including: gender, age, handedness and diagnostic information.

MCIC Dataset

Details about the MCIC dataset are described in a prior publication(Gollub et al., 2013).

Depression Dataset

Details about the MUC dataset are described in a prior publication(Koutsouleris et al., 2015).

1.10 External Validation Pipeline

The COBRE dataset has the highest duration of illness in years (M=16.8, SD=12.9), followed by MCIC (M=10.9, SD=10.9), and MUC (M=5.8, SD=7.7). The MUC dataset had the oldest subjects (M= 42.1, SD=11.9), followed by COBRE (M=38.1, SD=13.9), and MCIC (M=34.5, SD=11.1). Due to the significant age differences between the PRONIA recent onset samples and the older external validation samples we followed a robust pipeline that removed the age and site effects while retaining the disease variance in the data (see supplementary methods

1.10). Due to the significant age differences between the PRONIA recent onset samples and the older external validation samples HYDRA was not performed in the external datasets and we followed a robust pipeline that removed the age and site effects while retaining the disease variance in the data. First, we identified a HC subsample across different sites in the PRONIA and external validation datasets which was age and sex matched and had a specific age range (no older than 40 years of age). We then pooled the HC and patients from the PRONIA, COBRE, MCIC, and MUC datasets. Then, we trained ComBat on the identified HC subsample and applied it to the pooled HC and patients. Afterwards, we used the full ComBat corrected patient group and for each patient a normative ComBat corrected HC sample within an age window of ± 3 years was found. Then, the median and standard deviation was computed at the feature level and the patients' data were z-standardized. We then trained a repeated nested pooled cross-validation (CV) SVM model with 10 outer CV2 permutations, 10 outer CV2 folds, 10 inner CV1 permutations, and 10 inner CV1 folds using the z-standardized PRONIA data and applied the model to the z-standardized external validation datasets.

1.11 Cluster Significance Testing

Permutation testing is a robust tool for assessing statistical significance of machine learning models (Golland & Fischl, 2003; Nichols & Holmes, 2001). We assessed the statistical significance of clusters in three ways. First, we assessed whether the obtained subgroups were different to those which would be found in different disease structures by comparing the observed ARI to a null distribution of ARI created by randomly shuffling features in the HC group and the patient group and permuting 1000 times in HYDRA. Second, we assessed whether the obtained clusters were different to those which would be found if disease-related variability was not present by comparing the observed ARI to a null distribution of ARI created by randomly assigning HC samples ($n=275$) to a HC group and pseudo-patient group

and performing HYDRA analysis 1000 times. Third, we also assessed whether the data could be explained with clusters or if it could be better explained by a single Gaussian distribution by employing SigClust (1000 simulations) (Huang et al., 2015)

(<https://github.com/pkimes/sigclust2>). SigClust tests the null hypothesis that the data are from a single Gaussian distribution by using a test statistic called the cluster index (ClustI). The ClustI is defined as the sum of within-class sums of squares about the mean divided by the total sum of squares about the overall mean. The null distribution of the ClustI is approximated by simulating from a single Gaussian distribution estimated from the data.

Due to the unique nature of our recent onset samples we defined the null distribution by disturbing the disease signature in the patient group and the healthy control group. We performed random shuffling of features in the healthy control group and the patient group separately and permuted in HYDRA 1000 times using a repeated hold-out cross-validation strategy (i.e., 10 repetitions with 80% of the data for training in each repetition). We also permuted HYDRA 1000 times using the original data using a repeated hold-out cross-validation strategy (i.e., 10 repetitions with 80% of the data for training in each repetition). Thus, we obtained a null distribution ARI and compared it to the observed ARI. The ARI for the 2 cluster solution was statistically significantly higher in the observed distribution compared to the null distribution ($p_{\text{fdr}} < .001$). ARIs for all other observed clustering solutions showed no statistically significant differences to the null distribution ones ($p_{\text{fdr}} > 0.001$). We also performed permutation testing by randomly assigning HC samples ($n=275$) to a HC group and pseudo-patient group while maintaining the HC to patient ratio that was observed in the original data and performed HYDRA analysis 1000 times using a repeated hold-out cross-validation strategy (i.e., 10 repetitions with 80% of the data for training in each repetition)(Chand et al., 2020; Kaczkurkin et al., 2020). We compared the results to clustering results obtained with actual HC and patient groups. The ARI for the 2 cluster

solution was statistically significantly higher in the observed distribution compared to the null distribution ($p_{fdr}=0.01$).

1.12 SVM Details of Independent Validation and Prognostic Validation Models of 9-month GAF-S outcomes.

To further investigate the granularity of the differences between clusters, and manage the challenge of multiple testing in univariate statistics, we also trained Support Vector Machine (SVM) models to classify patients into the neuroanatomically based clusters using neurocognitive, blood-based biomarker, and symptom data and examined weighted features driving classification within the models. The machine learning analysis of pre-processed data was performed using NeuroMiner (version 1.0; <https://github.com/neurominer-git>). Prior to the machine learning analysis the following preprocessing steps were completed: a) every feature was scaled from 0 to 1 and completely non-finite features were zeroed-out; b) any feature that had infinite values was pruned; c) missing values were imputed using KNN Euclidean distance median replacement using 7 nearest neighbours. For each missing value of a given CV1 or CV2 subject, a subset of cases that had values for the given variable and had values in all other variables which were non-empty were identified. Subjects in the source subset were sorted according to their similarity with the target subject using the Euclidean distance. Then, the median of the given variable was computed using the 7 nearest neighbours. The original, non-imputed training matrix was used at all times. d) The data was scaled again. The model performance criterion that was used was balanced accuracy and the learning algorithm that was chosen was LIBLINEAR L2-regularized L2-loss SVC.

Imbalanced learning was corrected for by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes (weighting the hyperplane). The kernel type that was linear with eleven learning parameters in order to optimize the choice of C value. Wrapper methods were activated at all parameter combinations with greedy

sequential backward feature selection (Stop at $k=90\%$ of features; Feature stepping at 10% of worst performing features at each cycle). Starting with the full feature set the SVM was ran iteratively with the 10% worst performing features being eliminated at each cycle until 10% of the feature pool was left. In order to assess the models' statistical significance we performed permutation analysis to create an empirical null distribution of weights for each feature and then compare the observed weight to this distribution. The models were retrained in the cross-validation framework using the respective label subsets obtained from the observed-label analyses 1000 times. For each permutation the predictions were accumulated into a permuted ensemble prediction for each CV2 subject. In that way a null distribution of out-of-training classification performance for the prediction models was produced. The significance of the observed out-of-training classification performance was calculated as the number of events where the permuted out-of-training classification performance was higher or equal to the observed classification performance divided by the number of permutations performed. Then the significance of the model was determined according to a p threshold of $p<0.05$. Furthermore we applied a sign-based consistency algorithm to calculate the number of times that the sign of each feature (positive or negative) was consistent within an ensemble multiplied by the number of times that the feature was non-zero. The measure is between 0 to 1, with 1 representing perfect consistency within the ensemble and 0 if the weights are equally positive and negative or when the feature is omitted with a zero weight. A p -value was then calculated by defining a hypothesis test for the importance score with a null hypothesis of 0. A z -score was calculated as the importance divided by the square root of the variance of the importance scores. A standard p -value was then calculated using a normal cumulative distribution function to choose the right-tailed significance. P -values were then corrected using the false-discovery rate.

A repeated nested pooled cross-validation (CV) was used with 10 outer CV2 permutations, 10 outer CV2 folds, 10 inner CV1 permutations, and 10 inner CV1 folds. Imbalanced learning was corrected for, by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes. A linear kernel was used with 11 C values (0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16) to optimize the choice of C value and create an ensemble of predictive models to be applied to the CV2 data to produce a single average robust prediction. Balanced accuracy (BAC) regularized by SVM model complexity was used as a criterion for the hyperparameter optimization. For hyperparameter optimization, we computed

$$\overline{\text{BAC}}_{\text{reg}} = \sum_{i=1}^k \left(\frac{n_{TP_i}}{n_{TP_i} + n_{FN_i}} + \frac{n_{TN_i}}{n_{TN_i} + n_{FP_i}} \right) / 2$$

at given parameter combination across all k CV₁ partitions with $n_{TP_i}/(n_{TP_i} + n_{FN_i})$ being Sensitivity and $n_{TN_i}/(n_{TN_i} + n_{FP_i})$ being Specificity, and the fraction of the training population serving as support vectors in the i^{th} CV₁ partition. Our optimization technique's aim was to find a combination of T_G , PC s and the SVM's regularization parameter C [range: $2^{[-3 \rightarrow +4]}$] that maximized $\overline{\text{BAC}}_{\text{reg}}$ within a $3 (T_G) \times 5 (PC) \times 8 (C)$ hyperparameter cube. The optimized ensemble was then applied to the CV₂ validation data to produce a mean decision score ($\overline{D_{\text{ens}}}$) and majority voting-based class membership probabilities (P_{ens}) for each CV₂ validation subject. This produced a mean decision score ($\overline{(D_{\text{ens}})}$) and majority voting-based class membership probabilities (P_{ens}) for each CV₂ validation subject. Finally, a stacking-based data fusion framework (R. Polikar, 2006; Wolpert, 1992) was used to examine whether the combination of the symptom-based and the blood-based biomarker models would provide a superior classification accuracy. To achieve this, the decision scores of the blood-based biomarker and the symptom-based models were combined, standardized, and forwarded to a greedy sequential forward search algorithm (Saeys et al., 2007) which found a parsimonious combination of classifiers

maximizing $\overline{PSI_{reg}}$ across the C parameter range by employing L2-regularized logistic regression (L2LR)(Fan et al., 2008). Each L2LR ensemble was then applied to the standardized neuroimaging and clinical based decision scores available for the CV₂ validation data. Class prediction was achieved by using majority voting on Majority voting on P_{ens} .

1.13 Prognostic Validation Post-Hoc Results

Post-hoc Mann-Whitney U test results showed statistically significant differences between the BAC of the ROP stacked model and the Preserved Cluster stacked model ($U = .00$, $p < 0.001$), between the BAC of the ROD stacked model and the Preserved Cluster stacked model ($U = .00$, $p < 0.001$), and between the Impaired Cluster Stacked model and the Preserved Cluster stacked model ($U = .00$, $p < 0.001$), with no overlap of the values among the respective groups.

1.14 List of Features Included in Prognostic Validation Analysis

The symptom-based support vector machine learning model was trained using item scores the following tests: PANSS, BDI, and GAF-S. The blood-based biomarker model was trained using the following peripheral inflammatory marker data: IL1ra, CRP, TNF α , BDNF, and TGF β as well as the following additional variables: age, sex, tobacco use, and BMI.

1.15 Multivariate Analyses Results

Using symptom measures, and reflecting univariate results, the two clusters were separable (BAC: 60.9%, sensitivity: 65.4%, specificity: 56.4%, AUC: 0.63, $p < 0.01$). Assessing the feature weights of the model, patients in Cluster 2 exhibited higher scores in positive (PANSS-P items 6 and 7; suspiciousness/persecution and hostility), general (PANSS-G items 2 and 13; anxiety and disturbance of volition), negative symptoms (PANSS-N items 1 and 7; blunted affect and stereotyped thinking) and (SANS items 13, 14, 23, 25; alogia, grooming and hygiene, social inattentiveness, and attention) and BDI-II items 16 and 18 (changes in

appetite and sleeping patterns) among others. Patients in cluster 1 had higher scores in BDI-II items 4, 5, 8, 14, 15, and 17 (loss of pleasure, guilty feelings, self-criticalness, loss of energy, and irritability) and PANSS-G items 4, 6, 12, 14, and 15 (tension, depression, lack of judgement and insight, and poor impulse control) among others (See supplementary figure S3(a)).

Our neuroanatomically based clusters were also separable using cognitive data (BAC: 56.6%, sensitivity: 57.5%, specificity: 55.7%, AUC: 0.58, $p=0.01$). Patients in cluster 2 mainly exhibited worse cognitive performance in a visual recognition and recall task (Rey–Osterrieth complex figure) and patients in cluster 1 mainly performed worse in verbal memory tasks (Rey Auditory Verbal Learning Test) (See supplementary figure S3(b)).

The two clusters were also separable (BAC: 58.7%, sensitivity: 54.9%, specificity: 62.4%, AUC: 0.59, $p=0.01$) in blood-based biomarkers, with patients in cluster 2 having elevated levels of CRP and TNF α (See supplementary figures S3(c)).

Table S5: Clinical univariate comparisons between the two clusters (fdr-corrected p values)

	Cluster 1 (Preserved)	Cluster 2 (Impaired)	T value	P value
PANSS Positive M (SD)	11.5 (5.8)	13.8 (7.46)	-2.8	.02
PANSS Negative M (SD)	13.5 (6.3)	15.2 (13.5)	-2.2	.04
PANSS General M (SD)	29.8 (8.2)	33.0 (11.4)	-2.7	.01
BDI M (SD)	25.1 (14.2)	22.5 (12.7)	1.5	.12

Global Functioning Social Current M (SD)	6.0 (1.6)	5.7 (1.6)	1.5	.12
Global Functioning Role Current M (SD)	5.8 (1.8)	5.3 (1.8)	-2.3	.03
SANS Affective Flattening M (SD)	5.3 (7.6)	8.1 (9.1)	-2.7	.01
SANS Alogia M (SD)	2.1 (4.1)	3.8 (5.3)	-3.0	.02
SANS Avolition M (SD)	5.7 (4.6)	6.7 (5.3)	-1.7	.07
SANS Anhedonia M (SD)	8.8 (6.7)	9.8 (7.8)	-1.1	.25
SANS Attention M (SD)	1.8 (2.7)	2.8 (1.8)	-2.2	.04

**Table S6. Neurocognitive performance univariate comparisons between the two clusters
(fdr-corrected p values)**

	Cluster 1 (Preserved)	Cluster 2 (Impaired)	T value	P value
Verbal Fluency M (SD)	22.5 (6.7)	21.9 (7.0)	0.812	0.556
Social Cognition M (SD)	18.9 (2.5)	18.5 (2.7)	1.192	0.430
Working Memory M (SD)	15.6 (4.1)	15.5 (3.7)	0.121	0.904

TMA M (SD)	31.2 (13.3)	35.3 (16.9)	-2.212	0.074
DSST M (SD)	57.2 (12.6)	53.3 (13.9)	2.446	0.068
Reasoning M (SD)	19.6 (4.5)	19 (4.3)	1.107	0.430
Attention M (SD)	254.2 (24.6)	253.4 (20.1)	0.312	0.862
Speed of Processing M (SD)	48.5 (25.8)	40.0 (32.3)	2.399	0.068

Table S7. Blood-based biomarker univariate comparisons between the two clusters (fdr-corrected p values)

	Impaired Cluster	Preserved Cluster	T value	P value
IFN- γ M (SD)	4.2 (13.8)	2.8 (3.9)	-1.049	.296
IL1-ra M (SD)	615.4 (541.2)	802.3 (1304.6)	1.257	.210
IL-4 M (SD)	10.5 (10.8)	9.5 (8.6)	-.686	.493
S100B M (SD)	44.4 (61.1)	48.7 (82.4)	.409	.683
IL1- β M (SD)	.92 (.82)	3.0 (23.5)	.862	.390
IL-2 M (SD)	1.4 (1.9)	1.2 (1.7)	-.851	.396
IL-6 M (SD)	.83 (1.6)	.81 (.83)	-.072	.942

TNF- α M (SD)	1.6 (1.2)	1.6 (1.2)	.022	.983
CRP M (SD)	1562227.8 (4294394.8)	1380328.4 (1892965.3)	-.394	.694
TGF- β M (SD)	375230.8 (384248.7)	369705.3 (369923.8)	-.102	.919
BDNF M (SD)	21604.4 (9481.6)	22859.5 (9364.3)	.926	.356

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>Preserved Cluster > Impaired Cluster</i>		
Middle Frontal Gyrus, Superior Frontal Gyrus, Superior Temporal Gyrus, Medial Frontal Gyrus, Cingulate Gyrus, Right	0 0 -2	p(FDR)<0.001, T=8.43 4, k=201778

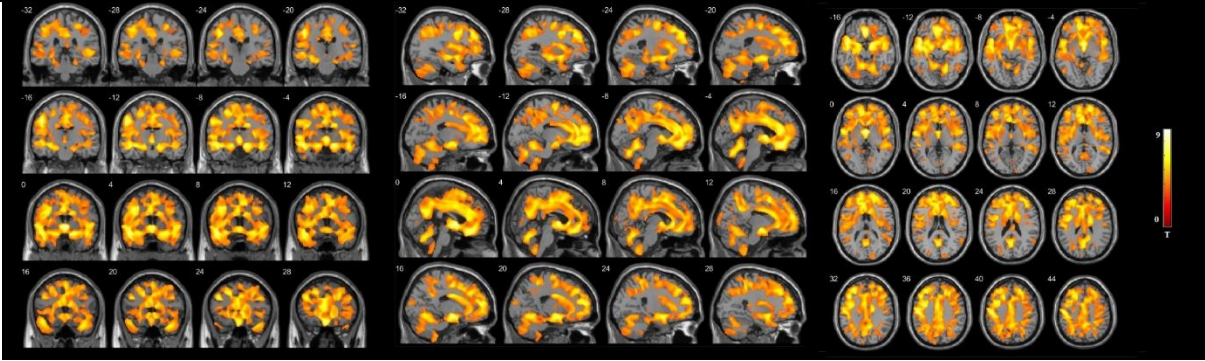
Cerebellum, Left Cerebellum, Precuneus, Precentral Gyrus, Inferior Frontal Gyrus, Anterior Cingulate, Insula, Parahippocampal Gyrus, Left Fusiform Gyrus, Hippocampus, Lingual Gyrus, Amygdala, Thalamus		
Cuneus, Middle Occipital Gyrus	18 -93 14	p(FDR)=0.001, T=4.28, k=645
Right Fusiform Gyrus, Inferior Temporal Gyrus, Middle Temporal Gyrus	47 -59 -14	p(FDR)=0.001, T=4.02, k=559

Table S7. Results of VBM Comparison between clusters

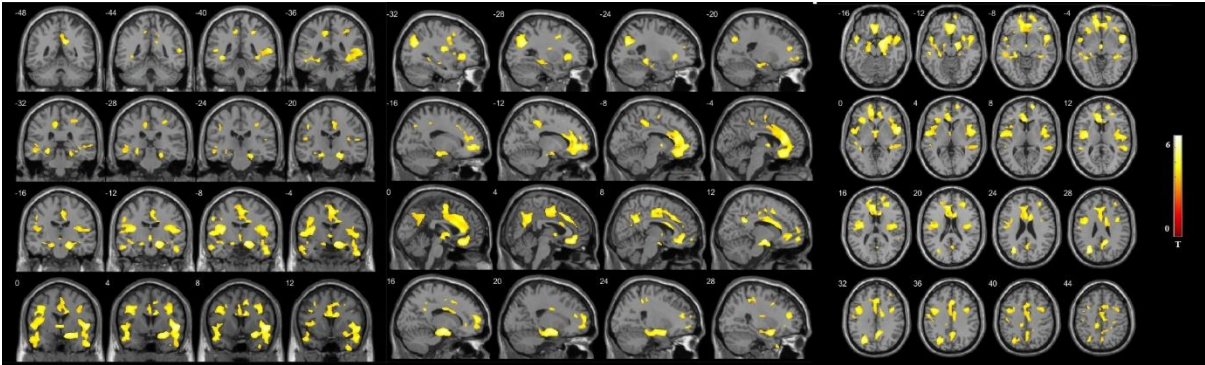
Table S8. Results of VBM comparison between clusters and HC

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>HC < Cluster 1</i>		
Bilateral Cerebellum, Left Lingual Gyrus, Fusiform Gyrus	-8 -54 -18	p(FDR)<0.001, T=4.01, k=1164
Bilateral Cerebellum	-2 -63 -30	p(FDR)=0.011, T=4.05, k=491
<i>HC > Cluster 2</i>		
Superior Temporal Gyrus, Right Insula, Right Superior Temporal Gyrus, Middle Temporal Gyrus, Parahippocampal Gyrus, Right Hippocampus, Inferior Frontal Gyrus, Right Amygdala, Fusiform Gyrus	53 6 -3	p(FDR)<0.001, T=5.46, k=9134
Anterior Cingulate, Medial Frontal Gyrus, Left Rectal Gyrus	-9 42 -3	p(FDR)<0.001, T=5.03, k=10983
Precuneus, Middle Temporal Gyrus,	-26 -68 27	p(FDR)<0.001, T=4.78, k=1376
Left Superior Temporal Gyrus, Left Insula, Left Inferior Frontal Gyrus, Left Hippocampus, Left Claustrum	-36 -9 14	p(FDR)<0.001, T=4.74, k=7684

Bilateral Precuneus, Cingulate Gyrus, Posterior Cingulate	11 -51 29	p(FDR)<0.001, T=4.54, k=1390
Superior Frontal Gyrus, Medial Frontal Gyrus	17 57 9	p(FDR)<0.001, T=4.36, k=1605
Bilateral Thalamus	0 -12 -5	p(FDR)=0.009, T=4.17, k=405
Right Postcentral Gyrus, Right Precentral Gyrus	26 -41 47	p(FDR)=0.010, T=4.12, k=386



(a)



(b)

Figure S2. GMV decreases between Impaired cluster and Preserved cluster (a) and between Impaired cluster and HC (b).

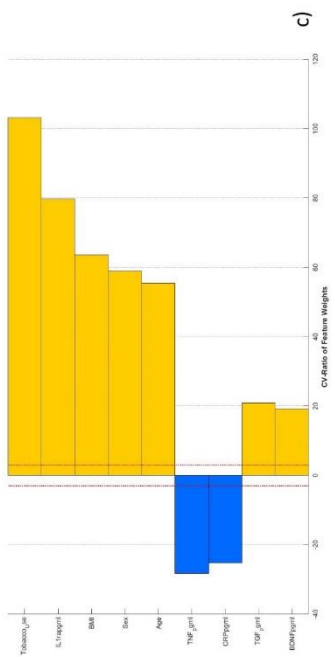
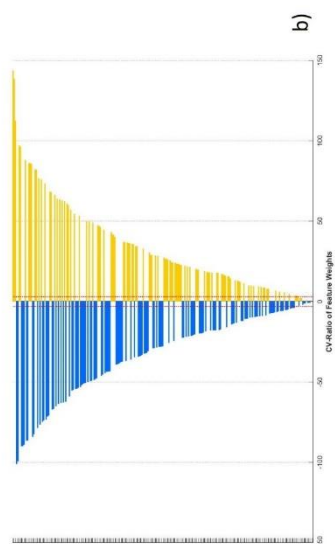
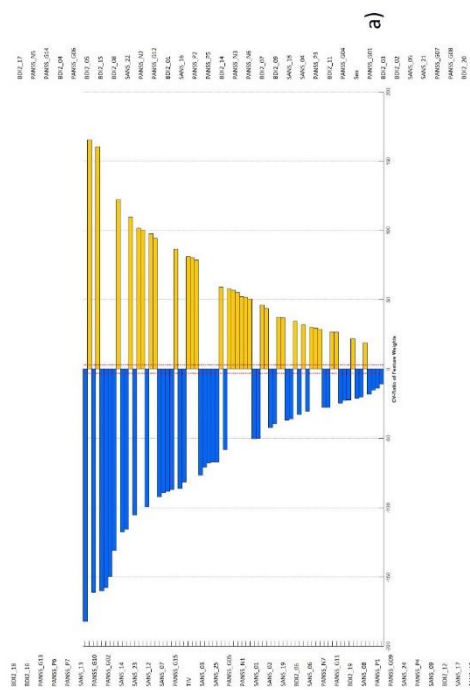


Figure S3. CV ratio of Feature Weights of SVM Classification between the Two Clusters

Using Symptom Measures (a), cognitive variables (b)*, and blood-based biomarker data (c).

Features in blue denote importance in impaired cluster classification and features in yellow denote importance in preserved cluster classification.

*The cognitive variables included in the model were (from top to bottom/most important to least important):

GAVLT_Delayed_repetition_out_of_list_words_list_A_T0

GROCF_Score_element_09_T0

GROCF_Score_element_05_Delayed_T0

GPVF_Correct_00_15_letter_1_T0

GFDS_Maxdigitsstringlengthcorrect_reminded_at_least_o_T0

GDSST_Score_symbol_matchings_T0

GDSST_Correct_number_symbol_matchings_T0

GROCF_Score_element_13_T0

GAVLT_Immedirepeatedwords1repetitionlistA_T0

GSVF_Correct_45_60_category_1_T0

SOPT_Maxcorrectresponsesbeforeerr10elem02_T0

GPVF_Repetition_15_30_letter_1_T0

GSVF_Error_45_60_category_1_T0

GROCF_Score_element_07_Immediate_T0

CPTIP_Number_error_filler_stimuli_200_trials_T0

SOPT_Perseveration_Errors_10_elements_03_T0

GROCF_Score_element_18_Delayed_T0

GROCF_Time_T0

CPTIP_Number_error_distracting_stimuli_50_trials_T0

Age

GAVLT_Immedioutoflist_words1repetitionlistA_T0

GPVF_Repetition_00_15_letter_1_T0

CPTIP_Reaction_times_error_distracting_300_trials_T0

GAVLT_Immedioutoflistwords5repetitionlistA_T0

GTMB_2_Errors_T0

GROCF_Score_element_06_T0

GSVF_Correct_30_45_category_1_T0

SOPT_Errors_10_elements_02_T0

GSVF_Repetition_30_45_category_1_T0

SOPT_Perseveration_Errors_6_elements_03_T0

GROCF_Score_element_16_Immediate_T0

GAVLT_Interference_Immediate_6_repetition_list_A_T0

GSVF_Repetition_00_60_category_1_T0

CPTIP_Number_correct_250_trials_T0

GAVLT_Immediate_5_repetition_list_A_T0

SOPT_Maxcorrectresponsesbeforeerr6elem01_T0

GAVLT_Immedirepeatedwords3repetitionlistA_T0

GTMA_1_Time_of_execution_T0

GROCF_Score_element_02_T0

GSVF_Repetition_15_30_category_1_T0

GROCF_Score_element_15_Immediate_T0

SOPT_Maxcorrectresponsesbeforeerr8elem01_T0

CPTIP_Number_error_filler_stimuli_100_trials_T0

GAVLT_Immedirepeatedwords2repetitionlistA_T0

GPVF_Error_45_60_letter_1_T0

CPTIP_Number_error_filler_stimuli_250_trials_T0

GROCF_Score_element_10_Immediate_T0

CPTIP_Reaction_times_error_distracting_100_trials_T0

GSVF_Correct_15_30_category_1_T0

GAVLT_Immedirepeatedwords4repetitionlistA_T0

GROCF_Score_element_14_Immediate_T0

GSAT_objects_rt_mean_T0

SOPT_Maxcorrectresponsesbeforeerr8elem03_T0

GROCF_Score_element_17_Immediate_T0

CPTIP_Number_error_distracting_stimuli_250_trials_T0

GPVF_Repetition_45_60_letter_1_T0

GROCF_Score_element_06_Delayed_T0

GTMA_3_Violations_T0

CPTIP_Number_error_distracting_stimuli_100_trials_T0

GSAT_animals_rt_sd_T0

GROCF_Score_element_02_Delayed_T0

GSAT_blue_objects_rt_mean_T0

GAVLT_Immediate_2_repetition_list_A_T0

GROCF_Score_element_14_T0

GROCF_Score_element_10_T0

CPTIP_Number_error_filler_stimuli_whole_test_T0

GROCF_Time_Delayed_T0

CPTIP_Reaction_times_error_distracting_200_trials_T0

CPTIP_Number_correct_200_trials_T0

GPVF_Correct_45_60_letter_1_T0

SOPT_Maxcorrectresponsesbeforeerr4elem02_T0

GROCF_Score_element_12_Delayed_T0

SOPT_Perseveration_Errors_4_elements_01_T0

GTMB_1_Time_of_execution_T0

CPTIP_Reaction_times_error_distracting_250_trials_T0

GROCF_Score_element_13_Delayed_T0

GFDS_Number_of_correct_trials_T0

SOPT_Maxcorrectresponsesbeforeerr10elem01_T0

GROCF_Placement_whole_Delayed_T0

GROCF_Score_element_04_T0

GROCF_Score_element_03_Delayed_T0

CPTIP_Reaction_times_error_distracting_150_trials_T0

GSAT_red_animals_vas_T0

CPTIP_Number_error_filler_stimuli_150_trials_T0

GTMB_3_Violations_T0

GROCF_Score_element_18_T0

GROCF_Score_element_08_Delayed_T0

WAIS_V_Standard_score_T0

GROCF_Score_element_11_Immediate_T0

WAIS_V_Raw_score_T0

SOPT_Errors_8_elements_02_T0

GROCF_Score_element_11_Delayed_T0

GSAT_red_objects_rt_mean_T0

GAVLT_InterfeImmedioutoflistwords6rep_lis_T0

CPTIP_Reaction_times_correct_50_trials_T0

GSVF_Error_00_15_category_1_T0

GROCF_Score_element_11_T0

GSAT_red_objects_rt_sd_T0

SOPT_Maxcorrectresponsesbeforeerr10elem03_T0

SOPT_Perseveration_Errors_8_elements_01_T0

SOPT_Maxcorrectresponsesbeforeerr6elem03_T0

GPVF_Error_00_60_letter_1_T0

GPVF_Error_15_30_letter_1_T0

GSVF_Error_00_60_category_1_T0

GSVF_Error_30_45_category_1_T0

GBDS_Maxdigitsstringlengthcorrect_reminded_at_least_o_T0

WAIS_MR_Standard_score_T0

SOPT_Perseveration_Errors_4_elements_03_T0

GROCF_Score_element_16_T0

SOPT_Maxcorrectresponsesbeforeerr8elem02_T0

GSAT_red_animals_rt_sd_T0

GSAT_blue_objects_rt_sd_T0

SOPT_Errors_10_elements_01_T0

GAVLT_Immedirepeatedwords5repetitionlistA_T0

GROCF_Score_element_07_T0

SOPT_Errors_6_elements_02_T0

Sex

GSAT_red_rt_mean_T0

CPTIP_Number_omissions_250_trials_T0

GROCF_Accuracy_whole_T0

GAVLT_InterfeImmediwordsfromlistArepetitionlist_T0

SOPT_Errors_6_elements_03_T0

GROCF_Score_element_04_Delayed_T0

GROCF_Score_element_17_T0

SOPT_Maxcorrectresponsesbeforeerr4elem03_T0

GAVLT_Immediate_4_repetition_list_A_T0

GROCF_Score_element_14_Delayed_T0

CPTIP_Number_omissions_300_trials_T0

GAVLT_InterfeImmedirepeatwordsrepetitionlistB_T0

GROCF_Score_element_15_T0

GROCF_Placement_whole_T0

GROCF_Score_element_08_Immediate_T0

GPVF_Correct_30_45_letter_1_T0

SOPT_Perseveration_Errors_6_elements_02_T0

CPTIP_Number_correct_300_trials_T0

GROCF_Score_element_03_Immediate_T0

SOPT_Errors_10_elements_03_T0

GPVF_Error_00_15_letter_1_T0

GAVLT_InterfeImmediwordsfromlistB6repetitionlis_T0

GAVLT_Immedioutoflistwords2repetitionlistA_T0

GROCF_Score_element_12_T0

CPTIP_Number_error_distracting_stimuli_150_trials_T0

GSAT_objects_rt_sd_T0

GROCF_Score_element_05_T0

GAVLT_Delayed_repetition_list_A_T0

GSAT_red_rt_sd_T0

GAVLT_InterfeImmedirepeatwords6repetitionlistA_T0

GROCF_Accuracy_whole_Immediate_T0

CPTIP_Number_omissions_50_trials_T0

GAVLT_InterfeImmedioutoflistwordsrepetitionlist_T0

GROCF_Score_element_07_Delayed_T0

SOPT_Errors_4_elements_01_T0

GPVF_Repetition_00_60_letter_1_T0

GPVF_Correct_00_60_letter_1_T0

GROCF_Score_element_10_Delayed_T0

GSAT_blue_objects_vas_T0

GROCF_Score_element_09_Delayed_T0

CPTIP_Number_correct_100_trials_T0

CPTIP_Reaction_times_correct_250_trials_T0

SOPT_Errors_8_elements_03_T0

GROCF_Score_element_09_Immediate_T0

SOPT_Errors_4_elements_03_T0

GPVF_Correct_15_30_letter_1_T0

CPTIP_Number_correct_whole_test_T0

GROCF_Score_whole_T0

CPTIP_Reaction_times_error_distracting_whole_test_T0

CPTIP_Reaction_times_correct_300_trials_T0

GSVF_Correct_00_15_category_1_T0

CPTIP_Number_omissions_150_trials_T0

GSAT_blue_rt_mean_T0

GPVF_Error_30_45_letter_1_T0

GSAT_blue_animals_rt_mean_T0

SOPT_Errors_4_elements_02_T0

SOPT_Errors_6_elements_01_T0

CPTIP_Number_error_distracting_stimuli_300_trials_T0

GSAT_animals_rt_mean_T0

GROCF_Score_element_08_T0

CPTIP_Reaction_times_correct_whole_test_T0

GROCF_Placement_whole_Immediate_T0

GROCF_Score_element_17_Delayed_T0

GROCF_Score_element_01_T0

SOPT_Maxcorrectresponsesbeforeerr6elem02_T0

CPTIP_Number_error_filler_stimuli_300_trials_T0

GAVLT_Delayed_repetition_words_from_list_B_list_A_T0

CPTIP_Number_omissions_200_trials_T0

SOPT_Maxcorrectresponsesbeforeerr4elem01_T0

GROCF_Score_element_05_Immediate_T0

GROCF_Score_element_01_Delayed_T0

GSVF_Error_15_30_category_1_T0

CPTIP_Reaction_times_correct_100_trials_T0

GROCF_Score_element_06_Immediate_T0

GSAT_total_rt_mean_T0

GROCF_Score_whole_Immediate_T0

GSAT_blue_animals_vas_T0

CPTIP_Number_error_distracting_stimuli_whole_test_T0

WAIS_MR_Raw_score_T0

CPTIP_Number_correct_150_trials_T0

GROCF_Score_element_13_Immediate_T0

GAVLT_Immedioutoflistwords4repetitionlistA_T0

GSVF_Correct_00_60_category_1_T0

CPTIP_Number_error_distracting_stimuli_200_trials_T0

SOPT_Perseveration_Errors_8_elements_02_T0

SOPT_Perseveration_Errors_8_elements_03_T0

CPTIP_Number_correct_50_trials_T0

GAVLT_Interference_Immediate_repetition_list_B_T0

SOPT_Perseveration_Errors_4_elements_02_T0

GROCF_Score_element_18_Immediate_T0

GSVF_Repetition_00_15_category_1_T0

GROCF_Score_element_12_Immediate_T0

GAVLT_Immediate_3_repetition_list_A_T0

GROCF_Score_element_15_Delayed_T0

CPTIP_Reaction_times_correct_200_trials_T0

CPTIP_Reaction_times_correct_150_trials_T0

CPTIP_Reaction_times_error_filler_whole_test_T0

CPTIP_Number_error_filler_stimuli_50_trials_T0

GROCF_Score_element_03_T0

GAVLT_Delayed_repetition_repeated_words_list_A_T0

GBDS_Number_of_correct_trials_T0

SOPT_Perseveration_Errors_6_elements_01_T0

GSAT_blue_animals_rt_sd_T0

GROCF_Score_element_04_Immediate_T0

SOPT_Perseveration_Errors_10_elements_01_T0

GSAT_blue_rt_sd_T0

GROCF_Accuracy_whole_Delayed_T0

GSAT_red_objects_vas_T0

GROCF_Time_Immediate_T0

GPVF_Repetition_30_45_letter_1_T0

GROCF_Score_element_16_Delayed_T0

GSAT_red_animals_rt_mean_T0

GDSST_Error_number_symbol_matchings_T0

GAVLT_Recognition_words_from_list_A_T0

CPTIP_Number_omissions_100_trials_T0

CPTIP_Number_omissions_whole_test_T0

GAVLT_Immediate_1_repetition_list_A_T0

GTMA_2_Errors_T0

GSAT_total_rt_sd_T0

SOPT_Perseveration_Errors_10_elements_02_T0

GAVLT_Immediate_1_repetition_list_A_T0

SOPT_Errors_8_elements_01_T0

CPTIP_Reaction_times_error_distracting_50_trials_T0

GDANVA_Number_correct_Faces_condition_T0

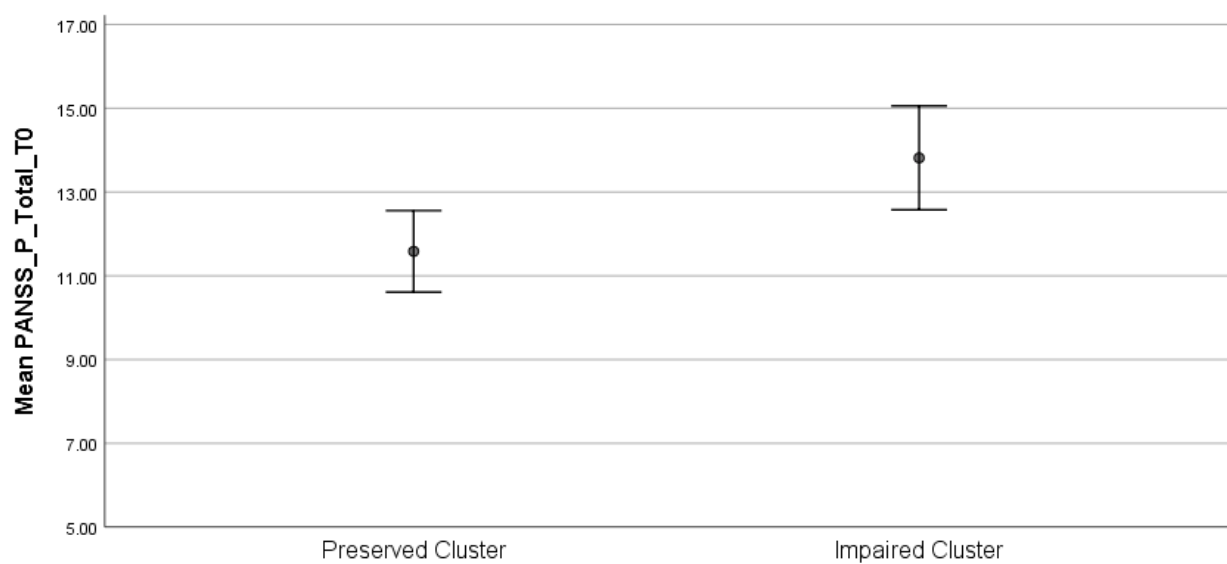
GROCF_Score_element_01_Immediate_T0

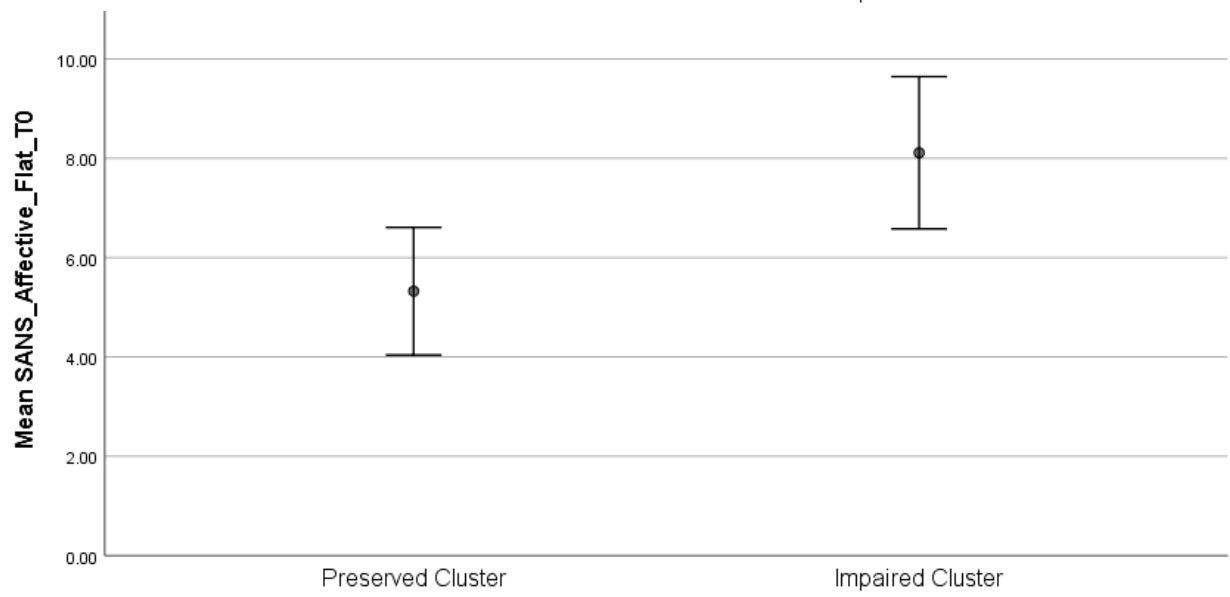
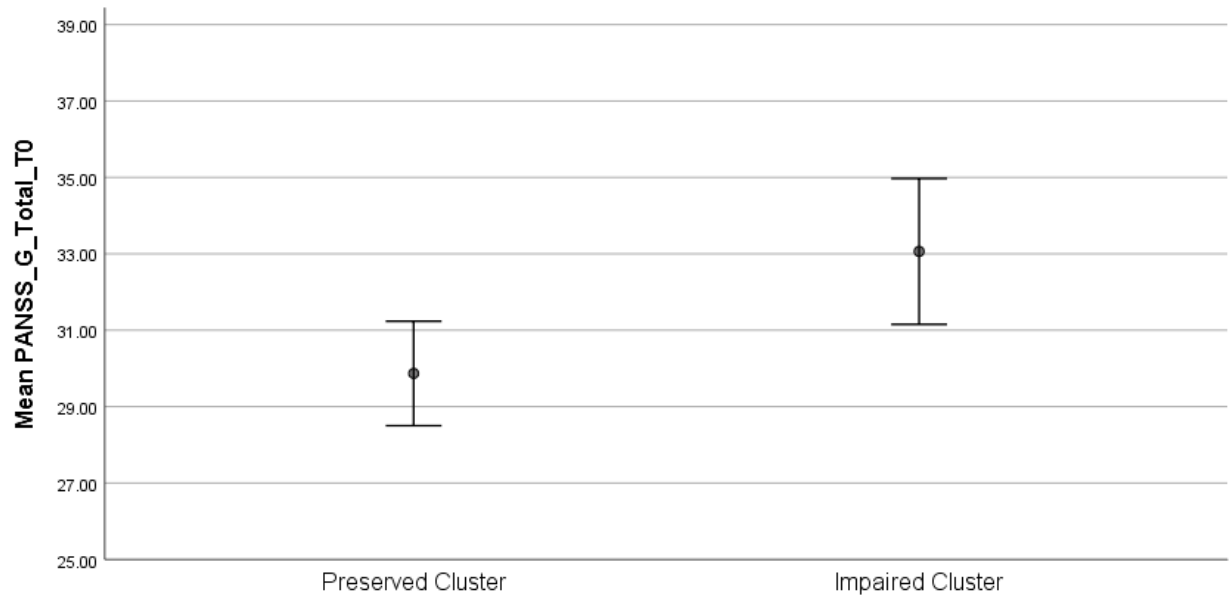
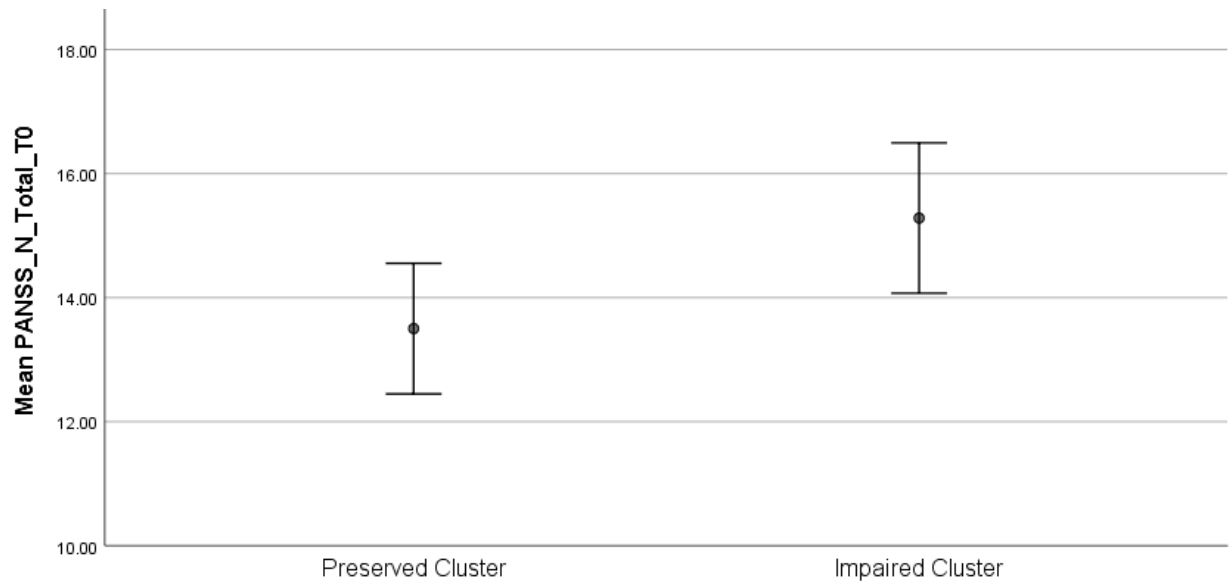
GROCF_Score_element_02_Immediate_T0

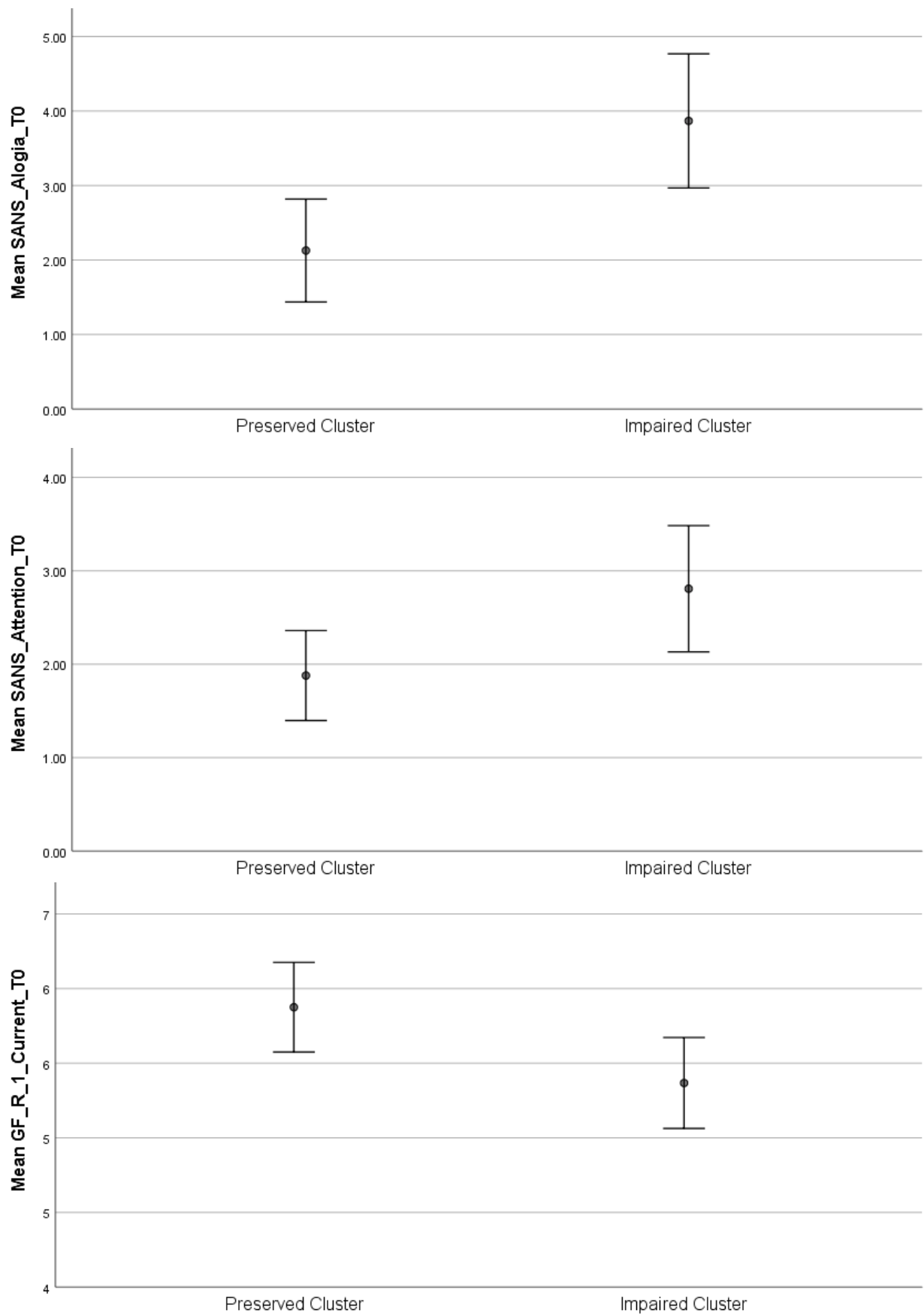
GSVF_Repetition_45_60_category_1_T0

GROCF_Score_whole_Delayed_T0

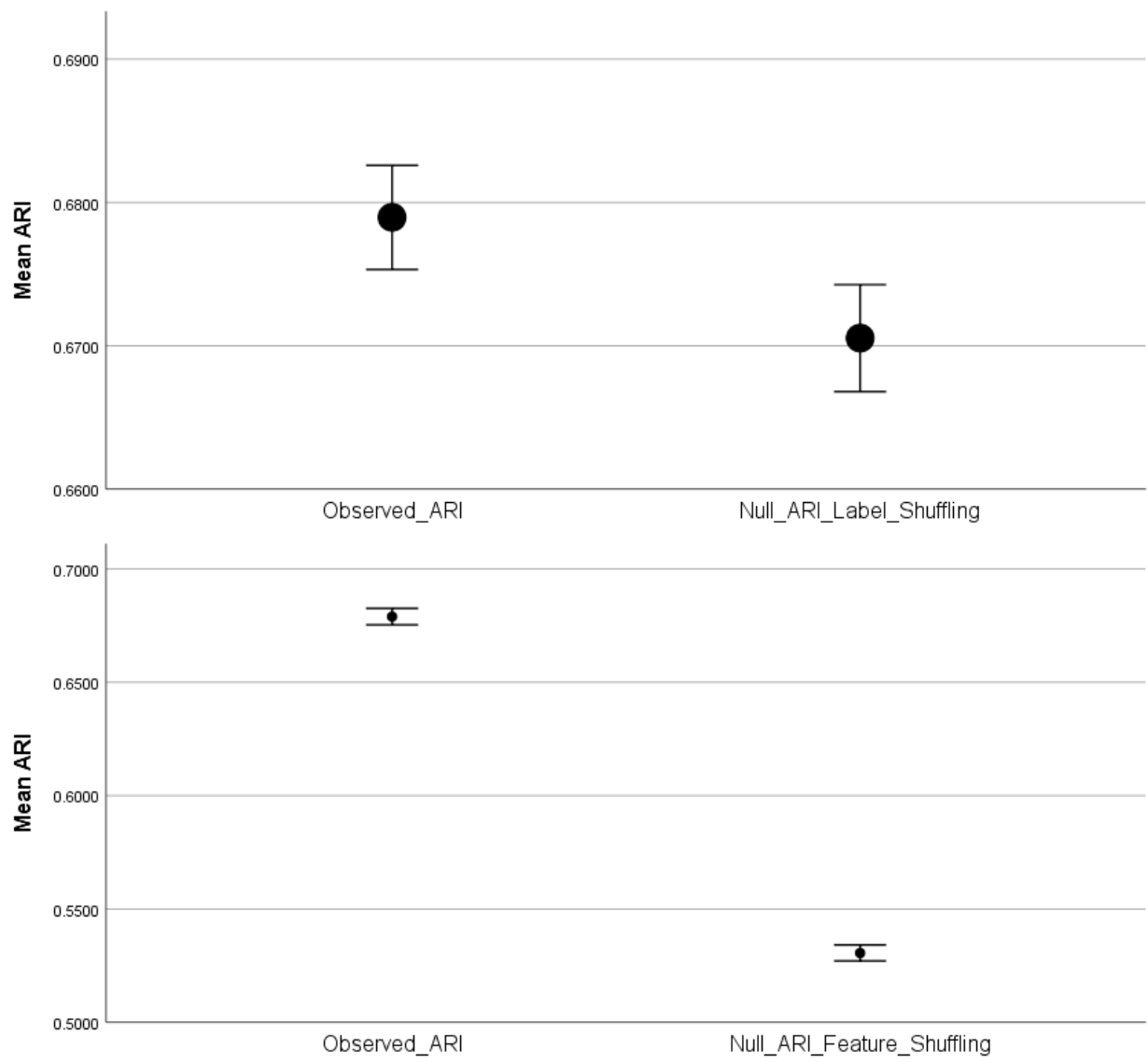
1.16 Error bar graphs of univariate analyses Discovery data



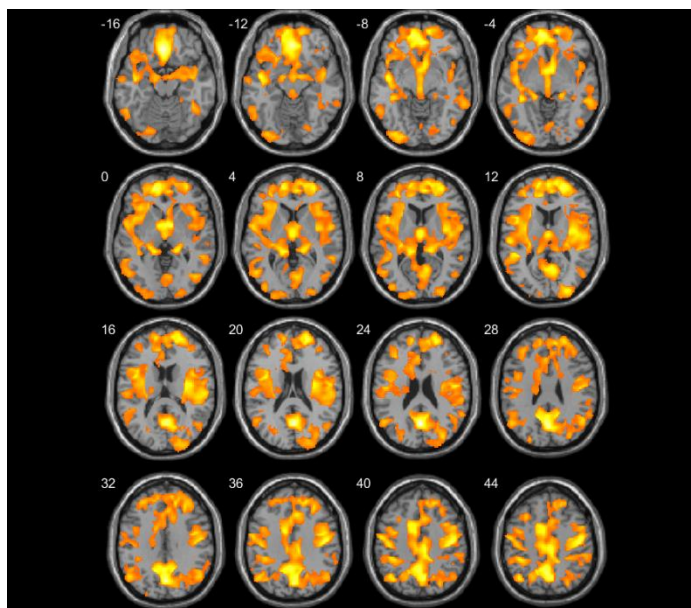
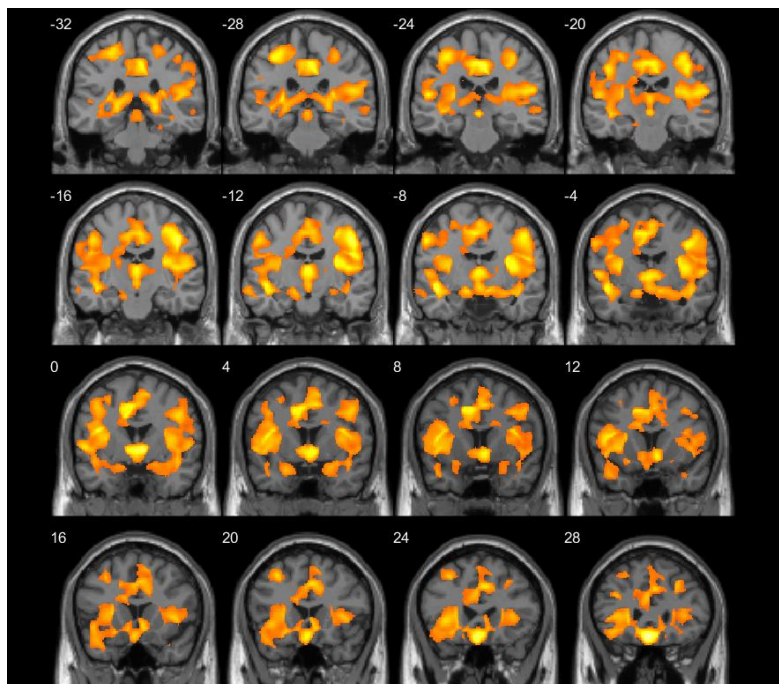


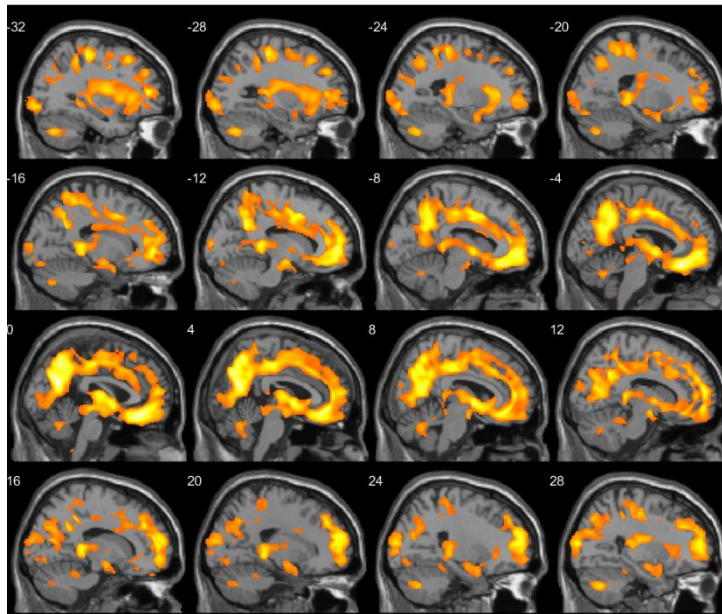


1.17 ARI significance testing error bars



1.18 Independent Validation VBM and Univariate analyses





cluster-level			peak-level					mm mm mm		
$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	(Z_E)	p_{uncorr}			
0.000	1252	0.000	0.000	0.000	8.11	7.57	0.000	2	-65	23
			0.000	0.000	7.56	7.12	0.000	-2	32	-17
			0.000	0.000	7.34	6.93	0.000	3	-72	35
0.000	2949	0.000	0.014	0.003	5.39	5.22	0.000	-35	-89	-9
			0.056	0.008	5.07	4.92	0.000	-26	-92	-9
			0.376	0.032	4.53	4.42	0.000	-39	-87	6
0.000	1434	0.000	0.063	0.008	5.04	4.89	0.000	54	-45	2
			0.102	0.011	4.91	4.78	0.000	59	-41	-5
			0.987	0.174	3.87	3.80	0.000	63	-26	-3
0.000	1850	0.000	0.075	0.009	4.99	4.85	0.000	29	-66	-38
			0.482	0.042	4.44	4.34	0.000	8	-63	-30
			0.663	0.060	4.29	4.20	0.000	-2	-62	-29
0.000	1362	0.000	0.094	0.010	4.93	4.80	0.000	-27	-68	-38
			1.000	0.392	3.54	3.49	0.000	-42	-56	-45
			1.000	0.772	3.25	3.21	0.001	-32	-54	-39
0.000	830	0.000	0.260	0.024	4.65	4.53	0.000	42	-50	-18
			0.977	0.155	3.92	3.85	0.000	39	-33	-12
			1.000	0.294	3.66	3.60	0.000	38	-38	-5
0.004	290	0.002	0.718	0.066	4.25	4.16	0.000	-62	-12	-14
0.004	307	0.001	0.981	0.161	3.90	3.83	0.000	20	-78	-11
0.003	326	0.001	0.996	0.205	3.80	3.73	0.000	14	-44	-23

Impaired<Preserved VBM Comparison

Cluster Peak Voxel MNI	
Coordinates (x,y,z)	
2 -65 23	Medial Frontal Gyrus, Precuneus, Cingulate Gyrus, Precentral Gyrus, Insula, Superior Frontal Gyrus, Middle Frontal Gyrus, Superior

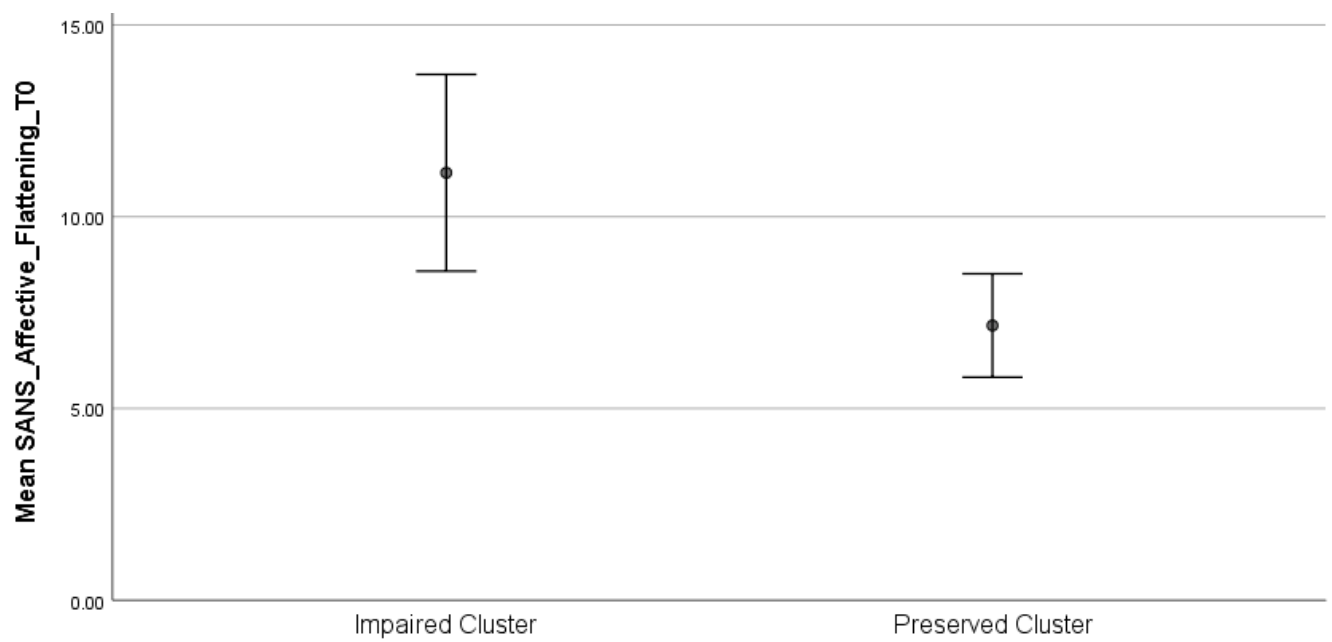
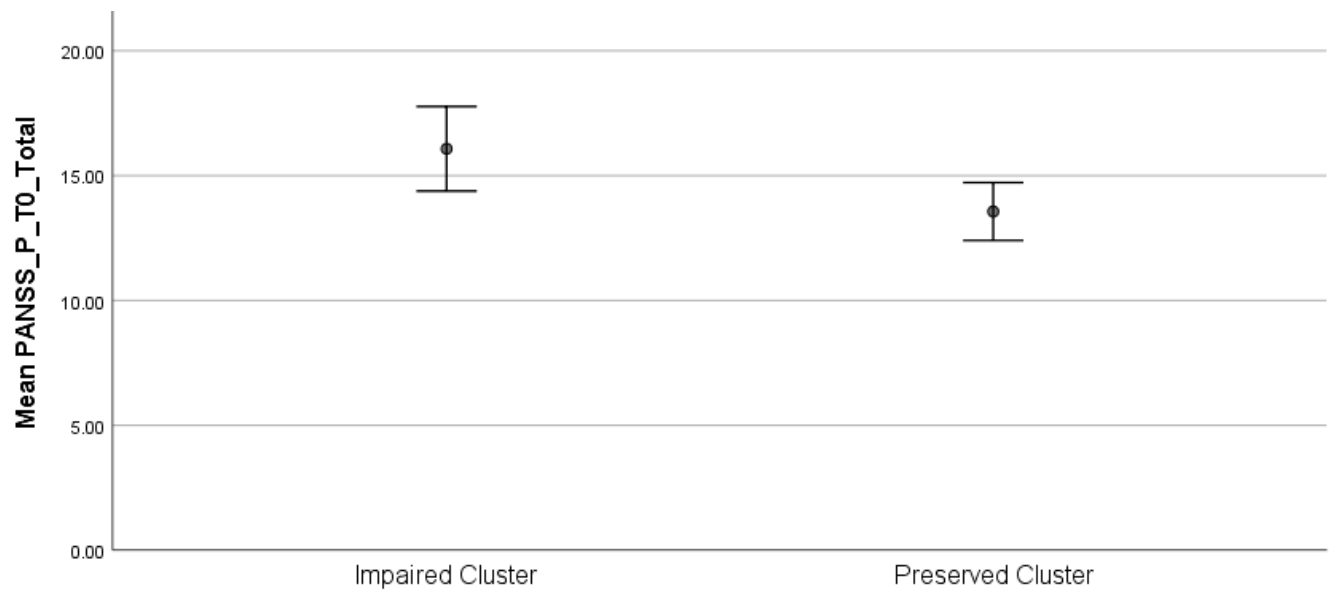
	Temporal Gyrus, Inferior Frontal Gyrus, Left Precuneus, Postcentral Gyrus, Right Precuneus, Anterior Cingulate, Thalamus, Parahippocampal Gyrus
-35 -89 -9	Middle Occipital Gyrus, Inferior Occipital Gyrus, Left Lingual Gyrus
54 -45 2	Middle Temporal Gyrus, Superior Temporal Gyrus
29 -66 -38	Right Cerebellum
-27 -68 -38	Left Cerebellum
42 -50 -18	Fusiform Gyrus
-62 -12 -14	Middle Temporal Gyrus
20 -78 -11	Right Lingual Gyrus
14 -44 -23	Right Cerebellum Anterior Lobe

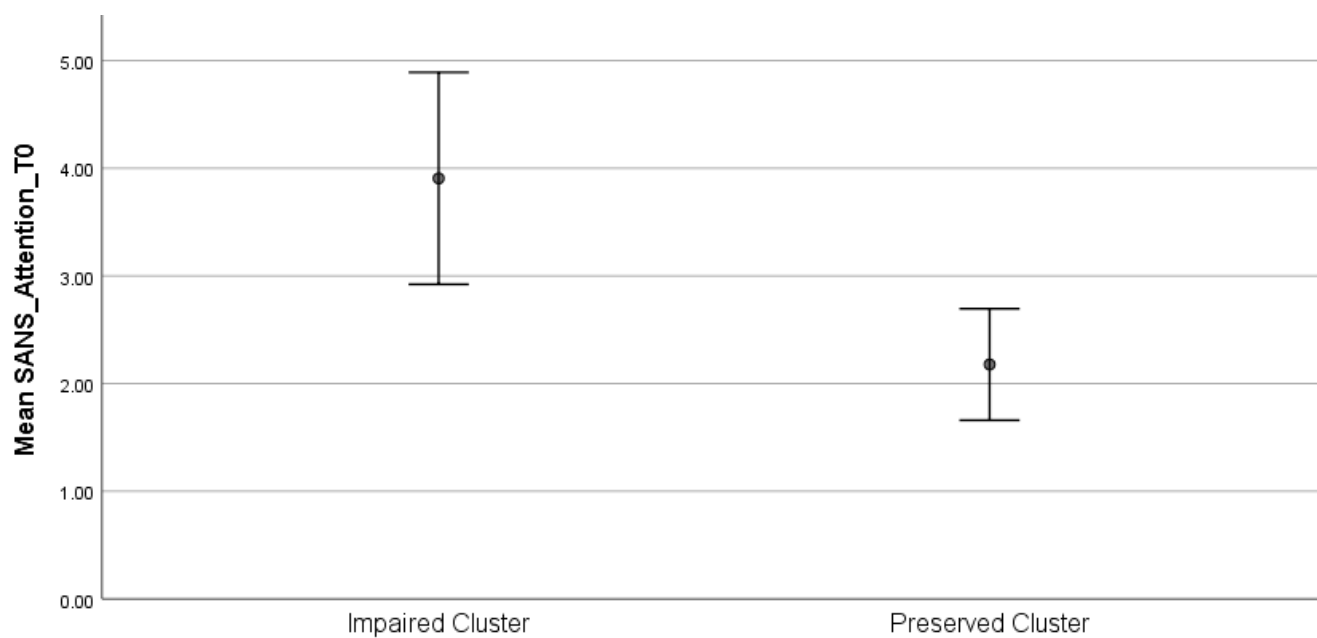
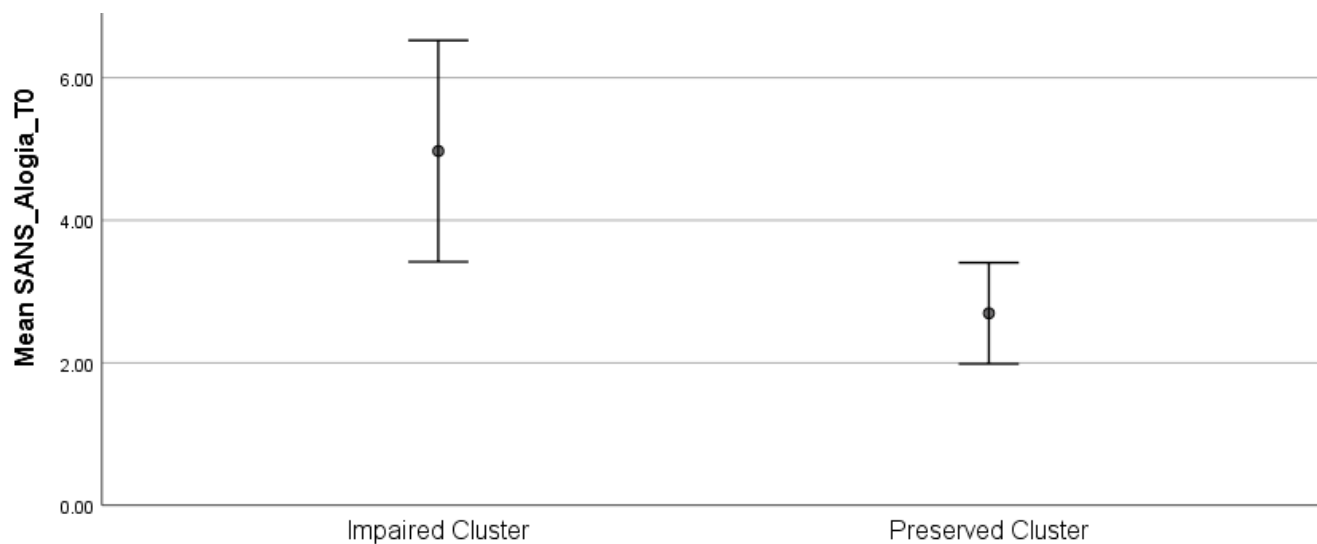
t-test for Equality of Means							
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
GF_S_1_Current_T0	-1.179	222	.240	-.263	.223	-.701	.176
GF_R_1_Current_T0	-.497	222	.620	-.138	.279	-.688	.411
IFN- γ (pg/ml)	.136	141	.892	.14891	1.09096	-2.00784	2.30566
IL-1ra (pg/ml)	.711	141	.478	57.61485	80.99184	-102.50047	217.73016
IL-4 (pg/ml)	-.528	141	.599	-.40099	.76014	-1.90373	1.10175
S100B (pg/ml)	.432	141	.667	4.82937	11.18814	-17.28882	26.94756
IL-1 β (pg/ml)	.702	141	.484	.14544	.20713	-.26405	.55493
IL-2 (pg/ml)	.369	141	.713	.14793	.40105	-.64491	.94078
IL-6 (pg/ml)	1.025	141	.307	.15718	.15336	-.14601	.46037
TNF- α (pg/ml)	-.221	141	.826	-.05614	.25430	-.55888	.44660
CRP (pg/ml)	1.707	141	.090	523068.11532	306451.25281	-82765.01609	1128901.24672
TGF - β (pg/ml)	1.013	141	.313	8729.78934	8621.03783	-8313.41182	25772.99049

BDNF (pg/ml)	.298	141	.766	281.49427	944.93808	-1586.58357	2149.57211
PANSS_P_T0_Total	2.452	213	.015**	2.51082	1.02398	.49240	4.52924
PANSS_N_T0_Total	1.586	212	.114	1.54603	.97489	-.37568	3.46774
PANSS_G_T0_Total	2.114	212	.036	2.95492	1.39794	.19928	5.71056
SANS_Affective_Flatte ning_T0	3.017	215	.003*	3.98277	1.31990	1.38117	6.58436
SANS_Alogia_T0	3.066	215	.002*	2.27507	.74191	.81271	3.73743
SANS_Avolition_T0	.510	215	.611	.39708	.77870	-1.13777	1.93194
SANS_Anhedonia_T0	.259	210	.796	.29330	1.13028	-1.93486	2.52146
SANS_Attention_T0	3.415	197	.001*	1.72847	.50614	.73031	2.72663

*False Discovery Rate corrected p values: 0.021

**False Discovery Rate corrected p value: 0.078





1.19 9 month remission between group comparison

		Remission		
		Remitted	Unremitted	Total
Cluster	Preserved	33	65	98
	Impaired	41	68	109
Total		74	133	207

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	.349 ^a	1	.555		
Continuity Correction ^b	.198	1	.656		
Likelihood Ratio	.349	1	.554		
Fisher's Exact Test				.565	.328
Linear-by-Linear Association	.347	1	.556		
N of Valid Cases	207				

Appendix D

Immune Subtypes of Schizophrenia and their Associated Brain Structure: a Semi-Supervised Machine Learning Examination of Inflammatory Heterogeneity.

Paris Alexandros Lalousis, MSc^{1,2}; Lianne Schmaal, PhD^{3,4}; Stephen J. Wood, PhD^{1,3,4}; Renate L.E.P Reniers, PhD^{1,2,5}; Vanessa Cropley, PhD⁶; Andrew Watson, PhD⁷; Christos Pantelis, MD, MRCPsych, FRANZCP⁶; John Suckling, PhD⁸; Nicholas M. Barnes, PhD, FBPhS⁹; Carmine Pariante, PhD¹⁰; Peter B. Jones, MD⁸; Eileen Joyce, PhD, MRCP, FRCPsych⁷; Thomas R.E. Barnes; DSc¹¹, Stephen M. Lawrie; MD, FRCPsych¹², Nusrat Husain, PhD¹³, Paola Dazzan, FRCPsych, PhD¹⁴; Bill Deakin, FRCPsych FMedSci¹⁵; Cynthia Shannon Weickert, PhD^{16,17,18}; Rachel Upthegrove, FRCPsych, PhD^{1,2,19}

Author Affiliations:

1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom

3 -Orygen, Parkville, Australia

4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

5 -Institute of Clinical Sciences, University of Birmingham, United Kingdom

6 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia

7 -The Department of Clinical and Motor Neuroscience, UCL Queen Square Institute of Neurology, London, UK

8 - Brain Mapping Unit, Department of Psychiatry, Herchel Smith Building for Brain and Mind Sciences, University of Cambridge, UK; Cambridgeshire & Peterborough NHS Foundation Trust, Cambridge, UK.

9 -Institute for Clinical Sciences, University of Birmingham, Birmingham, United Kingdom

10 - Stress, Psychiatry and Immunology Lab & Perinatal Psychiatry, The Maurice Wohl
Clinical Neuroscience Institute, King's College London, London, UK

11- Division of Psychiatry, Imperial College London, London UK

12- Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh,
Edinburgh, UK

13 - Division of Psychology and Mental Health, University of Manchester & Mersey Care
NHS Foundation Trust

14 -Department of Psychological Medicine, Institute of Psychiatry, Psychology and
Neuroscience, King's College London, London, UK

15 -Division of Neuroscience and Experimental Psychology, University of Manchester,
Manchester, United Kingdom

16 -Schizophrenia Research Institute, Darlinghurst, New South Wales, Australia

17 -Schizophrenia Research Laboratory, Neuroscience Research Australia, Sydney, New
South Wales, Australia

18 -School of Psychiatry, University of New South Wales, Sydney, New South Wales,
Australia

19 -Birmingham Early Interventions Service, Birmingham Women's and Children's NHS
Foundation Trust

Corresponding Author: Paris Alexandros Lalousis, BSc, MSc, School of Psychology,
University of Birmingham, 52 Pritchatts Road, Birmingham, B15 2SA, United Kingdom

1.1 External Dataset and Blood-Based Biomarker Data Storage and Analysis

We used data from 199 patients who had participated in the Benefit of Minocycline on Negative Symptoms of Psychosis: Extent and Mechanism (BeneMin) trial for exploration of our clusters in an external dataset. This is described in full in Deakin et al. (2018) but in brief, the participants had a mean age of 25.6 [S.D 5.2], were within 5 years of onset of symptoms judged sufficient to meet (DSM-IV) diagnostic criteria for a first episode of schizophrenia, schizophreniform, or schizoaffective psychosis, and 27% were female. Here, we use the demographic, clinical, cognitive and blood-based biomarker data from BeneMin acquired at entry to the trial.

ASRB: Serum samples were thawed at 4 °C and centrifuged at 1400g to remove any aggregate protein. The supernatant was then transferred to a fresh low-binding microfuge tube and diluted 1:2 in assay buffer according to the manufacturer's instructions. A 10-point standard curve with serial dilutions of 1:4 was generated using reconstituted stock standards supplied by the manufacturer; quality controls (QCs) supplied by the manufacturer were also used to determine assay accuracy. The data were generated against a 5-parameter logistic standard curve and corrected for background readings using Millipore Analyst Software (Merck Millipore, Billerica, MA, USA). Samples were run randomly across a total of eighteen 96-well plates in singles, and a pooled serum sample was used as an internal control on all plates. The average coefficient of variance for duplicate values (n = 160) across analytes was minimal (0.75%).

Serum was removed and stored frozen until sent to a central repository and inventoried, aliquoted, and stored at -80°C. Samples were sent to the Schizophrenia Research Lab in Sydney, where eight cytokines (IFN- γ , IL-10, IL-12(p70), IL-1 β , IL-2, IL-6, IL-8, and TNF α)

from the Human High Sensitivity T-Cell panel (HSTCMAG-28K, Merck Millipore) were assayed using a Luminex Magpix-based assay (Luminex corporation) and reported in picograms per millilitre (pg/ml). CRP was measured from serum by a high-sensitivity enzyme-linked immunosorbent assay (ELISA) according to the manufacturer's instructions (IBL-international, Hamburg, Germany).

BeneMin: Bloods for plasma inflammatory markers were aliquoted and frozen within 4 h. hs-CRP was measured with the Cormay anti-CRP antibody (PZ Cormay SA, Lomianki, Poland) sensitised to latex particles. Interleukin 6 (IL-6) was assayed with Meso Scale Discovery V-PLEX sandwich immunoassays (Rockville, MD, USA). Assays were done at King's College London.

1.2 MRI Data Acquisition and Processing

The T1-weighted (MPRAGE) images consisted of 176 1 mm thick sagittal slices per brain without gap; field of view = 250° - 250 mm; repetition time/echo time = 1980/4.3 ms; data matrix size = 256°-256; voxel dimensions = 1.0°-1.0°-1.0 mm. All scans underwent quality control and those with movement artefacts and/or brain lesions were excluded. The images were processed using the open-source CAT12 toolbox (version r1155; <http://dbm.neuro.uni.jena.de/cat12/>) (see supplementary methods 1.2). Whole-brain GMV measures were used.

The details of the MRI processing steps can be found in the CAT12 toolbox manual (<http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>) In short these steps consist of:

- (1) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering.
- (2) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local variations of intensity distributions as slowly varying spatial functions and thus achieves a homogeneous segmentation across cortical and subcortical structures.

- (3) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior information of adjacent voxels into the segmentation estimation generated by AMAP.
- (4) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM) inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP segmentation.
- (5) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the AMAP-generated tissue segments.
- (6) A high-dimensional DARTEL registration of the image to a MNI-template generated from the MRI data of 555 healthy controls in the IXI database (<http://www.braindevelopment.org>).
- (7) The GM maps were then multiplied with the Jacobian determinants that were obtained during registration in order to produce GM volume maps.
- (8) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

1.3 Statistical Significance

We created a null distribution of ARI by randomly assigning HC samples (n=600) to an HC group and pseudo-patient group and performing a HYDRA analysis 100 times using the aforementioned repeated hold-out cross-validation strategy.

SigClust tests the null hypothesis that the data are from a single Gaussian distribution by using a test statistic called the cluster index (ClustI). The ClustI is defined as the sum of within-class sums of squares about the mean divided by the total sum of squares about the overall mean.

The null distribution of the ClustI is approximated by simulating from a single Gaussian distribution estimated from the data.

1.4 Cluster Solution Generalizability SVM Details

We trained a repeated nested pooled cross-validation (CV) multi-class SVM model with 10 outer CV2 permutations, 10 outer CV2 folds, 10 inner CV1 permutations, and 10 inner CV1 folds to predict subgroup membership of patients in the ASRB sample in the clusters identified by HYDRA using the 9 cytokines/CRP. Secondly, this trained SVM model was applied to the BeneMin participants.

The machine learning analysis of pre-processed data was performed using NeuroMiner (version 1.0; <https://github.com/neurominer-git>). Prior to the machine learning analysis the following preprocessing steps were completed: a) every feature was scaled from 0 to 1 and completely non-finite features were zeroed-out; b) any feature that had infinite values was pruned; c) missing values were imputed using KNN Euclidean distance median replacement using 7 nearest neighbours. For each missing value of a given CV1 or CV2 subject, a subset of cases that had values for the given variable and had values in all other variables which were non-empty were identified. Subjects in the source subset were sorted according to their similarity with the target subject using the Euclidean distance. Then, the median of the given variable was computed using the 7 nearest neighbours. The original, non-imputed training matrix was used at all times. d) The data was scaled again. The model performance criterion that was used was balanced accuracy and the learning algorithm that was chosen was LIBLINEAR L2-regularized L2-loss SVC. Imbalanced learning was corrected for by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes (weighting the hyperplane). A linear kernel was used with 11 C values (0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16) to optimize the choice of

C value and create an ensemble of predictive models to be applied to the CV2 data to produce a single average robust prediction. Balanced accuracy (BAC) regularized by SVM model complexity was used as a criterion for the hyperparameter optimization. For hyperparameter optimization, we computed

$$\overline{\text{BAC}}_{\text{reg}} = \sum_{i=1}^k \left(\frac{n_{TP_i}}{n_{TP_i} + n_{FN_i}} + \frac{n_{TN_i}}{n_{TN_i} + n_{FP_i}} \right) / 2$$
 at given parameter combination across all k CV₁

partitions with $n_{TP_i}/n_{TP_i} + n_{FN_i}$ being Sensitivity and $n_{TN_i}/n_{TN_i} + n_{FP_i}$ being Specificity, and the fraction of the training population serving as support vectors in the i^{th} CV₁ partition. Our optimization technique's aim was to find a combination of T_G , PC s and the SVM's regularization parameter C [range: $2^{[-3 \rightarrow +4]}$] that maximized $\overline{\text{BAC}}_{\text{reg}}$ within a $3 (T_G) \times 5 (PC) \times 8 (C)$ hyperparameter cube. The optimized ensemble was then applied to the CV₂ validation data to produce a mean decision score ($\overline{D_{\text{ens}}}$) and majority voting-based class membership probabilities (P_{ens}) for each CV₂ validation subject. This produced a mean decision score ($\overline{(D_{\text{ens}})}$) and majority voting-based class membership probabilities (P_{ens}) for each CV₂ validation subject.. In order to assess the models' statistical significance we performed permutation analysis to create an empirical null distribution of weights for each feature and then compare the observed weight to this distribution. The models were retrained in the cross-validation framework using the respective label subsets obtained from the observed-label analyses 100 times. For each permutation the predictions were accumulated into a permuted ensemble prediction for each CV₂ subject. In that way a null distribution of out-of-training classification performance for the prediction models was produced. The significance of the observed out-of-training classification performance was calculated as the number of events where the permuted out-of-training classification performance was higher or equal to the observed classification performance divided by the number of permutations performed. Then the significance of the model was determined according to a p threshold of $p < 0.05$.

Furthermore we applied a sign-based consistency algorithm to calculate the number of times that the sign of each feature (positive or negative) was consistent within an ensemble multiplied by the number of times that the feature was non-zero. The measure is between 0 to 1, with 1 representing perfect consistency within the ensemble and 0 if the weights are equally positive and negative or when the feature is omitted with a zero weight. A p-value was then calculated by defining a hypothesis test for the importance score with a null hypothesis of 0. A z-score was calculated as the importance divided by the square root of the variance of the importance scores. A standard p-value was then calculated using a normal cumulative distribution function to choose the right-tailed significance. P-values were then corrected using the false-discovery rate.

1.5 Cluster Significance Testing

Permutation testing is a robust tool for assessing statistical significance of machine learning models(Golland & Fischl, 2003; Nichols & Holmes, 2001). We performed permutation testing by randomly assigning HC samples (n=275) to a HC group and pseudo-patient group while maintaining the HC to patient ratio that was observed in the original data and performed HYDRA analysis 1000 times using a repeated hold-out cross-validation strategy (i.e., 10 repetitions with 80% of the data for training in each repetition)(Chand et al., 2020; Kaczkurkin et al., 2020). We also permuted HYDRA 1000 times using the original data using a repeated hold-out cross-validation strategy (i.e., 10 repetitions with 80% of the data for training in each repetition) to obtain an observed ARI distribution. We compared the null distribution ARI results to the observed distribution ARI results. The ARI for the 8-cluster solution was statistically significantly higher in the observed distribution compared to the null distribution (p fdr <0.001). we also assessed whether the data could be explained with clusters or if it could be better explained by a single Gaussian distribution by employing SigClust (1000 simulations) (Huang et al., 2015) (<https://github.com/pkimes/sigclust2>). SigClust tests

the null hypothesis that the data are from a single Gaussian distribution by using a test statistic called the cluster index (ClustI). The ClustI is defined as the sum of within-class sums of squares about the mean divided by the total sum of squares about the overall mean. The null distribution of the ClustI is approximated by simulating from a single Gaussian distribution estimated from the data.

1.6 Further Inflammatory differences

The low inflammation cluster showed lower IL-10 levels (26.3 pg/ml (20.8) compared to cluster 7, $p<.001$); lower IL-12p70 levels (5.9 pg/ml (3.3) compared to HC, the CRP cluster, and clusters 3 and 7, $p<.001$); lower IL-2 levels (2.2 pg/ml (1.8) compared to HC, the CRP cluster, and clusters 3 ($p<.001$) and 7 ($p=.038$); lower IL-6 levels (52.5 pg/ml (185.7) compared to cluster 4, $p<.001$; and lower IL-8 levels (529.8 pg/ml (683.0)) compared to cluster 4, $p<.001$.

The classic inflammation cluster showed elevated levels of CRP (2.6 pg/ml (2.8) compared to HC ($p=.019$) and the low inflammation ($p=.006$) and IFN- γ ($p<.001$) clusters.

The IFN- γ showed elevated IL12p70 levels (14.9 pg/ml (12.5)) compared to HC and all other clusters apart from cluster 7, $p<.001$, and elevated IL-2 levels (6.5 pg/ml (5.2) compared to HC, the low inflammation cluster, and cluster 4, $p=.001$).

The anti-inflammatory cluster showed elevated levels of IL-2 (4.7 pg/ml (1.8) compared to the low inflammation cluster, $p=.038$.

1.6 Further VBM differences

The low inflammation cluster exhibited GMV decreases in areas such as the parahippocampal gyrus, inferior temporal gyrus, right hippocampus, middle temporal gyrus, fusiform gyrus, right amygdala, right inferior temporal gyrus, left hippocampus, and left amygdala when

compared to the HC. The anti-inflammatory cluster exhibited GMV decreases in the bilateral insula and the rectal gyrus when compared to the HC. Finally when all identified inflammatory clusters were compared with each other, decreases in the inferior temporal gyrus were observed in the CRP cluster when compared to the low inflammation and IFN- γ clusters; dorsolateral superior frontal gyrus GMV decreases were found in the CRP cluster when compared to the IFN- γ cluster; postcentral GMV decreases were seen in the CRP and classic inflammation clusters when compared to the IFN- γ cluster; and finally superior frontal gyrus GMV loss was observed in the classic inflammation cluster when compared to the IFN- γ cluster.

1.7 Neurocognitive Post-Hoc Comparisons

Post hoc analyses using the Tukey HSD post hoc criterion for significance indicated that patients in the IFN- γ cluster achieved greater scores on the RBANS language ($p=.007$) and attention ($p=.011$) tasks than patients in the combined low inflammation cluster and greater scores in the RBANS attention task compared with patients in the CRP ($p=.021$) cluster.

Patients in the IFN- γ cluster had higher scores in the LNS task and also had better premorbid functioning (WTAR) than patients in the combined low inflammation cluster ($p=.036$ and $.047$ respectively). No other differences in neurocognitive performance were observed (see supplementary Table S5).

1.8 Cluster Solution Generalizability SVM Performance

Patients in the CRP cluster were distinguishable from HC with a balanced accuracy (BAC) of 85.4% and area under the curve (AUC) of 0.93. Patients in the low inflammation cluster were separable from HC with a BAC of 68% (AUC: 0.77). Patients in the IFN- γ cluster could be separated from HC with a BAC of 69.6% (AUC: 0.72). BAC when separating HC from patients in the classic inflammation cluster was 84.7% (AUC 0.93) and 69.8 (AUC 0.70)

when separating them from patients in the anti-inflammatory cluster. Patients were distinguishable among different groups; CRP cluster-No Inflammation Cluster (BAC 87%, AUC 0.95); CRP cluster-IFN- γ cluster (BAC 95.4%, AUC 0.98); CRP Cluster-Classic Inflammation Cluster (BAC 89.9%, AUC 0.96); CRP Cluster-Anti-Inflammatory Cluster (BAC 90.4, AUC 0.97); Low Inflammation Cluster- IFN- γ cluster (BAC 93.7, AUC 0.98); Low Inflammation Cluster-Classic Inflammation Cluster (BAC 93.7, AUC 0.98); No Inflammation Cluster-Anti-Inflammatory Cluster (BAC 92.6, AUC 0.96); IFN- γ cluster-Classic Inflammation Cluster (BAC 96.3, AUC 0.98); IFN- γ cluster-Anti-Inflammatory Cluster (BAC 79.4, AUC 0.86); Classic Inflammation Cluster -Anti-Inflammatory Cluster (BAC 93.6%, AUC 0.96).

1.9 BeneMin Detected Clusters Inflammatory Details

Using the three clusters detected in the external BeneMin dataset (low inflammation cluster, CRP cluster, and anti-inflammatory cluster) and employing univariate ANOVAs, we found that there was a significant effect of cluster on cytokine expression which followed the patterns observed in the ASRB dataset, $F(2,196) = 101.01$, $p < .001$.

Post hoc analyses using the Tukey HSD post hoc criterion for significance indicated that mean CRP pg/ml levels were higher in the CRP cluster compared to the low inflammation cluster ($p < .001$) and the anti-inflammatory cluster ($p = .001$). Mean IFN- γ pg/ml levels were higher in the CRP cluster compared to the low inflammation cluster ($p = .003$) and lower compared to the anti-inflammatory cluster ($p = .001$). The low inflammation cluster exhibited lower IFN- γ pg/ml levels compared to the anti-inflammatory cluster ($p < .001$). IL-12 mean pg/ml levels were lower in the low inflammation cluster compared to the anti-inflammatory cluster ($p = .025$). IL-6 mean pg/ml levels were higher in the CRP cluster to the low inflammation cluster ($p < .001$). The CRP cluster also exhibited higher TNF α mean pg/ml

levels compared to the low inflammation cluster ($p=.002$) and lower compared to the anti-inflammatory cluster ($p=.038$). $\text{TNF}\alpha$ mean pg/ml levels were also lower in the low inflammation cluster compared to the anti-inflammatory cluster ($p=.003$). The $\text{IFN-}\gamma$ cluster was excluded from the ANOVA analysis due to having only one patient assigned to it. However, that patient had the highest $\text{IFN-}\gamma$ pg/ml levels (59.17) in the whole dataset.

1.10 BeneMin Detected Clusters Neurocognitive Details

When comparing the CRP cluster to the low inflammation cluster in neurocognitive domains, the patterns observed in the BeneMin dataset were similar to those found in the ASRB dataset, with patients in the CRP group exhibiting significantly lower current IQ ($p = .028$), lower premorbid IQ ($p = .005$), and worse performance on a processing speed task (WAIS digit symbol substitution task) ($p = .003$).

1.11 Supplementary Discussion

Studies of cytokines in schizophrenia have produced many inconsistent findings with cytokines reported as elevated in some studies but reported as normal in others (B. J. Miller & Goldsmith, 2019). This may be due to several factors, including the existence of inflammation subgroups that are variably sampled in different cohorts. Studies investigating adjunctive NSAID and tetracycline antibiotic treatment with anti-inflammatory properties in schizophrenia have so far produced heterogeneous results, with a potential explanation being that depending on the specific inflammatory subgroups enrolled, some cohorts may contain more potentially responsive people (inflamed people) than others (B. J. Miller & Goldsmith, 2019). If replicable inflammatory subgroups exhibiting differential inflammatory pathways can be demonstrated in schizophrenia, these could improve targeted novel treatments and help develop strategies of inflammatory stratification for the future design of randomized control trials (Khandaker et al., 2015).

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>HC > Classic Inflammation Cluster</i>		
Middle Frontal Gyrus, Inferior Frontal Gyrus, Superior Frontal Gyrus, Rectal Gyrus, Right Inferior frontal gyrus triangular part, Anterior Cingulate, Bilateral Insula	5 60 -21	p(FDR)<0.001, T=6.06, k=15435
Middle Temporal Gyrus, Superior Temporal Gyrus, Inferior Temporal Gyrus, Bilateral Hippocampus	63 -15 -27	p(FDR)<0.001, T=5.62, k=9519
Middle Temporal Gyrus, Inferior Temporal Gyrus, Fusiform Gyrus	-57 -24 -14	p(FDR)<0.001, T=6.06, k=5485
Superior Temporal Gyrus, Inferior Frontal Gyrus	-45 23 -23	p(FDR)<0.001, T=4.75, k=868
Left Insula	-44 3 0	p(FDR)<0.001, T=4.86, k=842
Superior Frontal Gyrus	-35 -92 2	p(FDR)=0.004, T=4.55, k=367
Inferior Parietal Lobule	-48 -45 45	p(FDR)<0.001, T=4.68, k=1568
Right Inferior Parietal Lobe	44 -47 56	p(FDR)=0.002, T=4.36, k=424

Table S1.

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>HC > CRP Cluster</i>		
Parahippocampal Gyrus, Inferior Temporal Gyrus, Right Hippocampus, Middle Temporal Gyrus, Fusiform Gyrus, Amygdala	32 -17 -18	p(FDR)<0.001, T=5.96, k=9726
Bilateral Thalamus	-14 -11 -30	p(FDR)<0.001, T=5.51, k=6027
Left Fusiform Gyrus	-62 -23 -26	p(FDR)<0.001, T=5.49, k=2410

Inferior Frontal Gyrus	-8 32 -24	p(FDR)<0.001, T=4.84, k=3029
Right Insula	54 17 -3	p(FDR)<0.001, T=4.74, k=1909
Left Lingual Gyrus	-5 -78 -11	p(FDR)<0.001, T=4.72, k=1278
Superior Temporal Gyrus	-41 20 -27	p(FDR)<0.001, T=4.39, k=768
Left Superior Frontal Gyrus	-26 56 9	p(FDR)<0.001, T=3.98, k=743
Left Insula	-39 17 0	p(FDR)<0.001, T=4.13, k=680
Inferior frontal gyrus, opercular part Right	48 8 42	p(FDR)<0.001, T=4.34, k=653
Right Superior Frontal Gyrus	30 51 6	p(FDR)<0.001, T=4.24, k=529
Right Lingual Gyrus	23 -53 -9	p(FDR)=0.001, T=4.13, k=476
Inferior Parietal Lobule, Angular Gyrus	45 -47 56	p(FDR)=0.001, T=4.38, k=459
Inferior frontal gyrus, orbital part Left	-32 38 -14	p(FDR)=0.002, T=4.42, k=389
Inferior frontal gyrus, opercular part Left	-41 11 27	p(FDR)=0.006, T=3.89, k=311

Table S2.

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>HC > Low Inflammation Cluster</i>		
Parahippocampal Gyrus, Inferior Temporal Gyrus, Right Hippocampus, Middle Temporal Gyrus, Fusiform Gyrus, Right Amygdala	30 -17 -18	p(FDR)<0.001, T=5.99, k=7143
Inferior Temporal Gyrus Right	51 3 -30	p(FDR)<0.001, T=4.24, k=1947
Left Hippocampus, Left Amygdala	-27 -20 -17	p(FDR)<0.001, T=4.85, k=2222
Inferior Frontal Gyrus	63 -26 -5	p(FDR)<0.001, T=4.87, k=1016
Left Lingual Gyrus	-5 -75 -11	p(FDR)=0.006, T=4.66, k=326
Middle Occipital Gyrus	-35 -92 2	p(FDR)=0.001, T=4.47, k=490
Right Superior Temporal Gyrus	48 -33 17	p(FDR)=0.003, T=4.39, k=393
Precuneus	-30 86 41	p(FDR)=0.002, T=4.36, k=424
Inferior Parietal Lobule	-47 -44 48	p(FDR)<0.001, T=4.93, k=1115
<i>HC < Low Inflammation Cluster</i>		
Right Pallidum	-18 -6 2	p(FDR)=0.052, T=4.79, k=140
Left Pallidum	21 -6 -2	p(FDR)=0.052, T=3.72, k=102

Table S3.

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>HC > IFN-γ Cluster</i>		
Parahippocampal Gyrus, Left Hippocampus	-22 -21 -15	p(FDR)=0.020, T=4.32, k=365
Parahippocampal Gyrus, Right Hippocampus	20 -15 -12	p(FDR)=0.023, T=3.76, k=251
Middle Temporal Gyrus	-62 -29 -2	p(FDR)=0.023, T=3.53, k=278
Inferior Parietal Lobule	-47 -39 50	p(FDR)=0.023, T=3.72, k=255
<i>HC < IFN-γ Cluster</i>		
Precentral Gyrus	-38 -21 72	p(FDR)<0.001, T=4.35, k=681

Table S4.

Comparison	Cluster Peak Voxel MNI Coordinates (x,y,z)	P Value, Peak Intensity, Cluster Size
<i>HC > Anti-Inflammatory Cluster</i>		
Right Insula	41 17 0	p(FDR)<0.001, T=5.33, k=2208
Rectal Gyrus	-11 26 -26	p(FDR)<0.001, T=4.34, k=660
Left Insula	-38 15 0	p(FDR)=0.026, T=4.25, k=233

Table S5.

	HC	Cluster 1 (Low Inflammation)	Cluster 2 (CRP Cluster)	Cluster 3 (Classic Inflammation Cluster)	Cluster 4 (IFN- γ Cluster)	Cluster 5 (Anti- Inflammatory Cluster)	ANOVA with HC		ANOVA Combined Clusters	P Value
							t/z/ χ^2 /F	P Value	F Value	P Value
Sample Sizes, No.	600	142	121	82	80	32				
Age, Mean (SD)	43.5 (13.1)	38.3 (11.1)	40.4 (10.7)	40.1 (10.8)	39.2 (11.4)	39.7 (9.5)	$F = .673$	.611		
Sex (Male/Female)	240/360	105/37	67/54	57/25	49/31	19/13	$\chi^2 = 11.5$	.021		
Freezer Time (Days, Mean (SD))	2427.9 (352.4)	2707.6 (412.3)	2497.0 (496.2)	2596.3 (464.6)	2632.6 (466.6)	2531.8 (383.8)	$F = 3.8$	.005		
Site No (Percentage)										
• Brisbane	• 193 (51.3)	• 44 (11.7)	• 49 (13.0)	• 42 (11.1)	• 34 (9.0)	• 11 (2.9)				
• Melbourne	• 178 (59.9)	• 40 (13.4)	• 28 (9.4)	• 19 (6.3)	• 14 (4.7)	• 11 (3.7)				
• Perth	• 97 (68.7)	• 17 (12.6)	• 12 (8.5)	• 6 (4.2)	• 7 (4.9)	• 2 (1.4)				
• Newcastle	• 55 (57.2)	• 16 (16.6)	• 11 (11.4)	• 1 (1.0)	• 8 (8.3)	• 5 (5.2)				
• Sydney	• 77 (49.0)	• 25 (15.9)	• 21 (13.3)	• 14 (8.9)	• 17 (10.8)	• 3 (1.9)	$\chi^2 = 21.2$	.168		
RBANS Immediate Memory	101.2 (.6)	78.6 (1.5)*	77.4 (1.7)*	78.3 (2.0)*	79.2 (2.1)*	77.3 (4.0)*	$F = 100.2$	>.001	$F = .205$	.893
RBANS Constructional	95.1 (.6)	85.4 (1.3)*	84.4 (1.5)*	82.3 (1.9)*	88.8 (1.6)*	84.1 (2.6)*	$F = 22.7$	>.001	$F = 2.01$	.112
RBANS Language	105.9 (.4)	91.8 (.9)*	94.7 (1.2)*	93.8 (1.3)*	96.7 (1.3)*	89.5 (2.7)*	$F = 56.7$	>.001	$F = 3.92$	.009
RBANS Attention	103.6 (.6)	80.2 (1.4)*	80.2 (1.6)*	81.8 (1.8)*	87.0 (2.1)*	77.4 (2.7)*	$F = 98.6$	>.001	$F = 3.73$	.011
RBANS Delayed Memory	98.7 (.4)	81.8 (1.4)*	81.9 (1.6)*	77.9 (2.1)*	80.8 (1.9)*	83.6 (3.0)*	$F = 81.3$	>.001	$F = 1.07$	.361
RBANS Total Score	100.9 (.5)	78.8 (1.2)*	79.0 (1.3)*	77.9 (1.6)*	82.5 (1.6)*	77.4 (2.9)*	$F = 138.6$	>.001	$F = 1.63$	.180
WTAR Standard Score	103.8 (.4)	96.5 (1.2)*	93.9 (1.4)*	94.4 (1.7)*	99.5 (1.5)	94.8 (3.7)*	$F = 22.0$	>.001	$F = 2.76$	.041
WASI Vocabulary	58.1 (.3)	47.5 (.9)*	47.9 (1.0)*	48.7 (1.3)*	50.7 (1.3)*	46.1 (2.5)*	$F = 52.2$	>.001	$F = 1.60$	.188
WASI Matrix Reasoning	58.9 (.3)	50.6 (.9)*	49.7 (.9)*	50.5 (1.3)*	53.9 (1.1)*	49.3 (2.3)*	$F = 50.1$	>.001	$F = 2.73$	.043
WASI IQ Equivalent	115.0 (.4)	99.2 (1.2)*	98.7 (1.4)*	100.1 (1.8)*	104.5 (1.7)*	97.6 (3.3)*	$F = 70.0$	>.001	$F = 2.92$	.033
LNS Score	11.9 (.1)	8.6 (.2)*	8.9 (.2)*	9.2 (.3)*	9.7 (.3)*	8.2 (.6)*	$F = 54.1$	>.001	$F = 2.87$	.036

Table S6. Sample Sociodemographics. Sample Sizes, Participants per Study Site, Age, Sex, Freezer Time, Neurocognitive Performance (Higher values represent better neurocognitive performance, F values represent main effect of group on neurocognitive performance, asterisks represent tests which are statistically significantly lower to HC). Abbreviations: SD=Standard Deviation, RBANS=Repeatable Battery for the Assessment of Neuropsychological Status, WTAR=Weschler Test of Adult Reading, WASI=Weschler Abbreviated Scale of Intelligence, IQ=Intelligence Quotient, LNS=Letter-Number Sequencing).

	Low Inflammation Cluster			CRP cluster		Classic
	N	Mean	Std. Deviation	Mean	Std. Deviation	Mean
sans_affect_global_rating	117	1.60	1.443	1.27	1.297	1.59
sans_alogia_global_rating	118	1.24	1.482	.89	1.218	1.11
sans_asociality_global_rating	119	2.19	1.520	2.26	1.639	2.12
gaf_scale_response	120	51.36	13.319	49.22	12.861	51.61
sans-avolition-global_rating	119	1.84	1.519	2.09	1.535	2.02

Table S7. SANS subscales and Global Assessment of Functioning Scores

	(Low Inflammation Cluster)	(CRP Cluster)	(IFN-γ Cluster)	(Anti-Inflammatory Cluster)	t/z/ χ ² /F
Sample Sizes, No.	149	48	1	2	
Age, Mean (SD)	26.9 (5.1)	26.9 (5.1)	39.2 (11.4)	39.7 (9.5)	F = 1.0
Sex (Male/Female)	109/40	35/13	0/1	2/0	χ ² = 1.0
CRP (pg/ml) (Mean (SD))	1.59 (1.3)	9.39 (6.3)	.80 (1.3)	.25 (.21)	F = 1.0
IFN-γ (pg/ml) (Mean (SD))	3.99 (3.1)	7.79 (12.7)	59.1 (3.1)	25.3 (8.5)	F = 1.0
IL-10 (pg/ml) (Mean (SD))	.38 (.44)	.36 (.26)	.59 (.44)	.65 (.09)	F = 1.0
IL-12 (pg/ml) (Mean (SD))	.12 (.09)	.15 (.25)	.12 (.09)	.41 (.48)	F = 1.0
IL-1β (pg/ml) (Mean (SD))	.07 (.24)	.06 (.13)	.01 (.24)	.01 (.00)	F = 1.0
IL-2 (pg/ml) (Mean (SD))	.26 (.25)	.29 (.26)	.97 (.25)	.41 (.07)	F = 1.0

IL-6 (pg/ml) (Mean (SD))	.64 (.41)	1.16 (.75)	.94 (.41)	.52 (.22)	<i>F</i> =
IL-8 (pg/ml) (Mean (SD))	4.44 (3.0)	4.41 (2.5)	11.9 (3.0)	2.94 (.16)	<i>F</i> =
TNFα (pg/ml) (Mean (SD))	2.44 (.66)	2.79 (.48)	3.3 (.66)	3.91 (.81)	<i>F</i> =

Table S8. BeneMin Clusters Inflammatory Characteristics

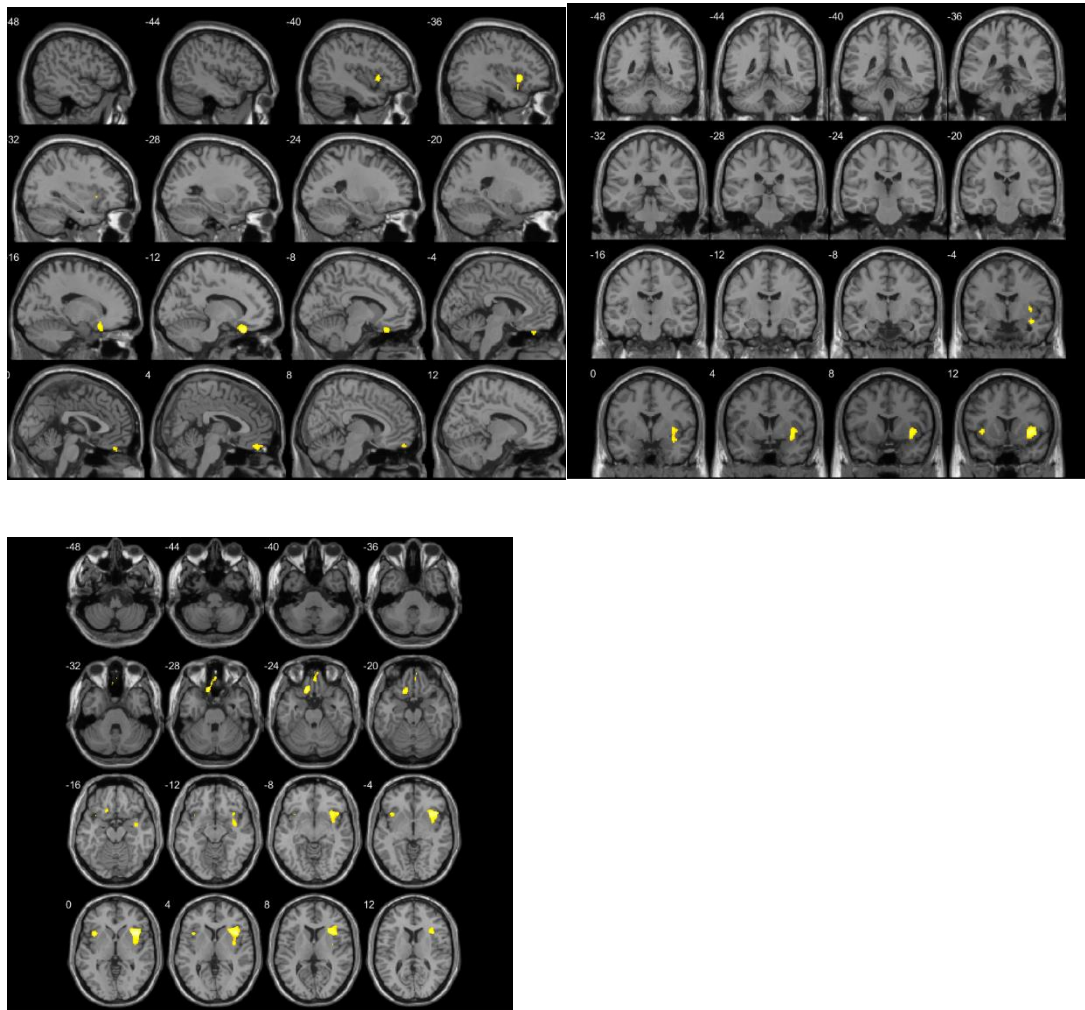


Figure S1.